

Quantum thermodynamics with constrained fluctuations in work

Jonathan G. Richens^{1,2} and Lluís Masanes²

¹*Controlled Quantum Dynamics theory group, Department of Physics,
Imperial College London, London SW7 2AZ, UK.*

²*Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK.*

In the standard framework of thermodynamics the work produced or consumed in a process is a random variable whose average is bounded by the change in the free energy of the system. This average work is calculated without regard for the size of its fluctuations. We find that in some processes, such as reversible cooling, the fluctuations of the work given by the change in free energy can diverge. Small or fragile thermal machines may be unable to cope with arbitrarily large fluctuations. Hence, with the present focus on nano-scale thermodynamics, we analyse how thermodynamic efficiency rates are modified when the size of these fluctuations are restricted. We quantify the work content and work of formation of arbitrary finite dimensional quantum states when the work fluctuations are bounded by a given amount c . By varying c we interpolate between the work given by the change in the standard free energy $c=\infty$ and the min-free energy $c=0$, defined in the context of single-shot thermodynamics. We derive fundamental relations between average and deterministic work, and explore the emergence of irreversibility and partial order on state transformations when bounding fluctuations. We also study the efficiency a single qubit thermal engine model with constrained fluctuations, and derive the corrections to the Carnot efficiency.

Historically, thermodynamics has been a theory of macroscopic systems comprising of many particles. As we venture away from the thermodynamic limit we must question the validity of established principles. Recently, the problem of extracting work from a microscopic quantum system has received much attention [1–5]. The standard free energy is used to calculate the maximal amount of average work that can be extracted from a system in thermal contact with an infinite heat bath. Generally the work extracted on each running of the protocol fluctuates, but in the thermodynamic limit the relative size of fluctuations in work vanishes. However, in the case of microscopic systems, and systems that are far from equilibrium, fluctuations in the work can no longer be ignored. It is of significant practical importance that we understand these fluctuations in order to describe the behaviour of small and fragile machines such as quantum heat engines comprising of just a few qubits [6–8]. Realistic thermal machines are designed to operate at specific energies with a certain tolerance to fluctuations. Taking into account this inevitable *fragility* requires a modified free energy that tells us the average work associated with a process when fluctuations in that work are constrained.

One approach to dealing with fluctuations is to simply not allow for them. This is the tactic employed by single-shot thermodynamics, a recently developed approach to quantum thermodynamics inspired by the field of single-shot information theory [1, 4, 5]. The single-shot (deterministic) work associated with a process is given by the difference in min-free energy between initial and final states [1, 4], which is generally significantly smaller than the standard free energy difference. The work cost of forming a state from the Gibbs state is given by the max-free energy [1, 4], which is generally significantly larger than the deterministic work that can be extracted from the state. This discrepancy between the work cost and work content of states in the single-shot regime results in *thermodynamic irreversibility* when transforming between states. Furthermore, the set of allowed thermodynamic transformations in the single-shot regime are severely restricted. In this regime, it is possible for a state to undergo a transition $\rho \rightarrow \rho'$ deterministically

(and without supplying work) on the condition that an infinite family of “second laws” are satisfied [9]. Some transitions $\rho \leftrightarrow \rho'$ can only be achieved by supplying work in both forward and backward directions, resulting in a *partial order* on the set of states with respect to the resource of work [9]. This is in stark contrast to when we allow work to fluctuate freely, whereby all states can be inter-converted in a thermodynamically reversible manner. One may be tempted to assume that the fundamental irreversibility and partial order are idiosyncratic of the single-shot regime, due to the strong constraint of requiring work to be a deterministic quantity. In this article we show that there are always state transformations that can only be achieved in a thermodynamically irreversible way, or exhibiting a partial order with respect to work, if we demand only that work cannot fluctuate infinitely.

To date the majority of thermodynamic protocols treat work either as an unconstrained random variable or a totally constrained (deterministic) quantity. In this article we explore the landscape of protocols that exist between these two regimes. We find that many processes, for example thermodynamically reversibly cooling, result in diverging fluctuations in work. This makes realising these protocols practically infeasible, especially for small and fragile machines. To this end we define the *c-bounded work*, giving the optimal average work $\langle w \rangle$ associated with a process for which all fluctuations of the random variable w are bounded as

$$|w - \langle w \rangle| \leq c \quad (1)$$

where c is a adjustable parameter. We explore how bounding work fluctuations affects work extraction, state formation and the allowed state transformations of individual systems. Finally, we use the *c-bounded work* to derive a corrected Carnot efficiency for a single qubit engine model when fluctuations in the work produced by the engine are constrained.

The organisation of the paper is as follows. In the first section we introduce the framework within which we derive our results. In the second and third sections we discuss deterministic and unbounded average work

extraction and work of formation, and discuss the conditions under which fluctuations associated with reversible processes diverge. We then introduce and characterise the c -bounded work, and in the final section we apply these results to the study of a single qubit heat engine.

The framework. In this section we provide a precise description of our framework, describing the system, bath, work system and the set of allowed operations.

We make use the widely applied set-up for thermodynamic protocols of system, infinite thermal bath and a weight, which acts as a store and source of the work produced or consumed by a protocol [2, 3, 10]. In the following we set the Boltzmann constant k_B to 1. The bath has infinite volume and it is in the Gibbs state $\rho_B = \frac{1}{\mathcal{Z}_B} e^{-\beta H_B}$, where β is the inverse temperature, H_B the Hamiltonian, and \mathcal{Z}_B the partition function.

The work system is modelled as a suspended weight with a continuous energy spectrum and Hamiltonian dependent only on its displacement $H_W = \int_{\mathbb{R}} dx x |x\rangle\langle x|$, where the orthonormal basis $\{|x\rangle, \forall x \in \mathbb{R}\}$ represents the position of the weight. In order to define work as a classical random variable w , the position of the weight is measured at the beginning and end of the protocol.

The system being transformed has Hilbert space of dimension d , initial state and Hamiltonian (ρ, H_S) , and final state and Hamiltonian (ρ', H'_S) . It is useful to define the initial and final dephased states and their spectral decompositions

$$\int dt e^{-iH_S t} \rho e^{iH_S t} = \sum_s x_s |s\rangle\langle s|, \quad (2)$$

$$\int dt e^{-iH'_S t} \rho' e^{iH'_S t} = \sum_{s'} x_{s'} |s'\rangle\langle s'|. \quad (3)$$

The two bases $|s\rangle$ and $|s'\rangle$ defined above, allow to write the spectral decompositions $H_S = \sum_s \mathcal{E}_s |s\rangle\langle s|$ and $H'_S = \sum_{s'} \mathcal{E}'_{s'} |s'\rangle\langle s'|$. (Note that we use notation $x_{s'}$ and $\mathcal{E}'_{s'}$ instead of $x'_{s'}$ and $\mathcal{E}'_{s'}$.) Finally, we assume that initially the joint state of system, bath and weight is product $\rho \otimes \rho_B \otimes \rho_W$. We consider any process that is a joint transformation of system, bath and weight represented by a Completely Positive Trace Preserving (CPTP) map Γ_{SBW} satisfying the following conditions:

Microscopic reversibility (Second Law): It has an (CPTP) inverse Γ_{SBW}^{-1} , which implies unitarity $\Gamma_{\text{SBW}}(\rho_{\text{SBW}}) = U \rho_{\text{SBW}} U^\dagger$.

Energy conservation (First Law): $[U, H_S + H_B + H_W] = 0$.

Independence from the “position” of the weight: The unitary commutes with the translations on the weight $[U, \Delta_W] = 0$. The generator of the translations Δ_W is canonically conjugated to the position of the weight $[H_W, \Delta_W] = i$.

Classicality of work: Before and after applying the global map Γ_{SBW} the position of the weight is measured, obtaining outcomes $|x\rangle$ and $|x+w\rangle$ respectively. The joint transformation of the system and

work random variable w is given by the map

$$\Lambda(\rho_S, w) = \int_{\mathbb{R}} dx \text{tr}_{\text{BW}} [Q_{x+w} U (\rho_S \rho_B Q_x \rho_W Q_x) U^\dagger], \quad (4)$$

where $Q_x = |x\rangle\langle x|$ is a weight position projector.

The assumption that the dynamics of a closed system is reversible and conserves energy is widely used, because it corresponds to a common physical setup. The third condition implies that the reduced map on the system and bath is a mixture of unitaries, and therefore cannot decrease the entropy of the joint state of system and bath (See Result 1 in [10]). This ensures that the weight cannot be used as a source of non-equilibrium, and can be viewed as a necessary condition for defining work [2, 11]. A consequence of the fourth condition is that the dynamics of the “diagonal part” of the state $x_s \rightarrow x_{s'}$ decouples from the rest of the transformation $\rho \rightarrow \rho'$ (See Appendix 1C). In the case of optimal work extraction and work of formation we find that the initial / final state is diagonal (See Appendix 3A). Therefore we need only to consider to diagonal part of the state, denoted x_s , when determining the work content and work of formation of a state.

Deterministic work. The single-shot work content of a system is given by the difference in min free energy $F^{\min}(\rho) = -\beta^{-1} \log \sum_s x_s^0 e^{-\beta \mathcal{E}_s}$ between the state ρ and the thermal state

$$W^{(0)}(\rho) = \frac{1}{\beta} \log \mathcal{Z} - \frac{1}{\beta} \log \sum_s x_s^0 e^{-\beta \mathcal{E}_s}, \quad (5)$$

where x^0 returns 1 if $x > 0$ and 0 if $x = 0$, and $\mathcal{Z} = \text{tr}(e^{-\beta H_S})$ is the partition function of the system [1, 4]. If x_s has full rank then min free energy is $-\beta^{-1} \log \mathcal{Z}$. Therefore non-zero deterministic work can only be extracted from states that are not of full rank. The single-shot work of formation is

$$W_F^{(0)}(\rho) = \frac{1}{\beta} \log \mathcal{Z} + \frac{1}{\beta} \log \max_s x_s e^{\beta \mathcal{E}_s}. \quad (6)$$

In general we find that $W_F^{(0)}(\rho) > W^{(0)}(\rho)$, i.e. it is not possible to form most states in a thermodynamically reversible manner. When a weight is not present, the necessary and sufficient condition for a state transformation $(\rho, H_S) \rightarrow (\rho', H'_S)$ to be possible is given by the thermo-majorisation criteria [1]. If in addition a catalyst is used, the necessary and sufficient conditions are given in [9]. The key phenomenon is that in single-shot thermodynamics there is a partial order on states, i.e. there are state transitions that are impossible, in both forward and backward directions, without supplying work. In contrast thermodynamic irreversibility and partial order are not observed in the standard thermodynamic formalism, in which work is allowed to fluctuate freely, as we discuss next.

Work with unbounded fluctuations. The maximum average work that can be extracted from a system with the assistance of a heat bath with inverse temperature β is given by the difference in free energy between ρ and the Gibbs state. The free energy is given by

$F(\rho) = \langle \mathcal{E} \rangle - \beta^{-1} S(\rho)$, $\langle \mathcal{E} \rangle = \text{tr}(\rho H_S) = \sum_s x_s \mathcal{E}_s$ is the internal energy of the state, $S(x_s) = -\sum_s x_s \log x_s$ is the entropy of the de-phased state. The Gibbs state, with energy level occupation probabilities $x_s = \mathcal{Z}^{-1} e^{-\beta \mathcal{E}_s}$, is the unique state given H_S and β with the lowest free energy (given by $-\beta^{-1} \log \mathcal{Z}$). Therefore the optimal average work that can be extracted from an out of equilibrium state is given by $W^{(\infty)} = \beta^{-1} \log \mathcal{Z} + F(\rho)$. In the reverse process of state formation the work cost is also given by the difference in free energy between initial and final states. In other words, if we do not bound fluctuations in the work it is always possible to realise all state transformations in a thermodynamically reversible way. This poses the question - what is the minimum amount we must allow work to fluctuate in order for a transition to be achievable with a thermodynamically reversible protocol?

Result 1. *There exists a thermodynamically reversible process achieving the transition $(\rho, H_S) \rightarrow (\rho', H'_S)$ with fluctuations in work less than or equal to c if*

$$e^{\beta(\Delta F - c)} \leq \frac{x_s e^{\beta \mathcal{E}_s}}{x_{s'} e^{\beta \mathcal{E}_{s'}}} \leq e^{\beta(\Delta F + c)} \quad \forall s, s', \quad (7)$$

where $\Delta F = F(\rho) - F(\rho')$ is the change in the standard free energy. This becomes a necessary and sufficient condition if the initial and/or final state is diagonal in the energy eigenbasis

Proof. See Appendix 2 □

Note that for any finite c there exist states such that (7) cannot be satisfied. These bounds have strong consequences for the minimal fluctuations that can be achieved with a thermodynamically reversible protocol. For example

$$\lim_{x_i \rightarrow 0} \log \frac{x_s e^{\beta \mathcal{E}_s}}{x_{s'} e^{\beta \mathcal{E}_{s'}}} = - \lim_{x_{s'} \rightarrow 0} \log \frac{x_s e^{\beta \mathcal{E}_s}}{x_{s'} e^{\beta \mathcal{E}_{s'}}} = -\infty \quad (8)$$

Therefore as an energy level occupation probability tends to zero the work fluctuation associated with transitioning to or from this energy level diverges, negatively for work extraction and positively for state formation, when performing a thermodynamically reversible protocol. In either case we require $c \rightarrow \infty$ in order to satisfy inequalities (7).

Result 1 tells us that the further from equilibrium the initial or final states are, the larger the work fluctuations will be in thermodynamically reversible protocols. Note that cooling a system close to its ground state is an example of transitioning from the thermal state to a far from equilibrium state. Similarly, if we want to extract work from a far from equilibrium state using a thermodynamically reversible transformation we encounter the same divergence in fluctuations. The fluctuations can diverge even if the average work remains small (for example, if the system is a qubit with trivial Hamiltonian then $W \leq \beta^{-1} \log 2$). These divergences have been previously noted in the recent study of *absolute irreversibility* [12, 13].

Previous discussions of the inadequacy of the standard free energy in the nano-regime have focused on the definitions of work [4, 5]. Here we add another criticism,

that using the standard free energy to describe work we necessarily requires set-ups that can tolerate arbitrary fluctuations, which diverge in size for processes with initial or final states that are increasingly far from equilibrium.

Work with bounded fluctuations.

Motivated by these observations we define the *c-bounded work content* $W^{(c)}(\rho)$ as the maximum average work that can be extracted from state ρ with initial Hamiltonian H_S and final Hamiltonian H'_S , when the fluctuations of the work are constrained by c , as in (1). Analogously, we define the *c-bounded work of formation* $W_F^{(c)}(\rho)$ as the minimal average work that is necessary to create a state ρ with Hamiltonian H'_S from the Gibbs state (with respect to initial Hamiltonian H_S), such that fluctuations in the work are bounded by c .

Result 2. *The c-bounded work content $W^{(c)}(\rho)$ and work of formation $W_F^{(c)}(\rho)$ are given by*

$$W^{(c)}(\rho) = \frac{1}{\beta} \log \mathcal{Z}' - \frac{1}{\beta} \log \sum_s x_s^0 e^{-\beta(\mathcal{E}_s - \theta_s^{(c)})} \quad (9)$$

$$W_F^{(c)}(\rho) = \frac{1}{X_u} \left[\sum_{s \in \mathcal{X}_u} \frac{x_s}{\beta} \log(x_s e^{\beta \mathcal{E}_s} \mathcal{Z}) + c(1 - X_u) \right] \quad (10)$$

Proof. See Appendices 3 and 4 respectively □

First we describe the terms in (9). $\mathcal{Z}' = \sum_{s'} e^{-\beta \mathcal{E}_{s'}}$ is the partition function of the final Hamiltonian H'_S . The second term can be viewed as a generalisation of the min free energy (5) that allows for fluctuations in work $\theta_s^{(c)}$. The only difference to the min free energy is the term $e^{\beta \theta_s^{(c)}}$ included in the summation. $\theta_s^{(c)}$ is the fluctuation of the work marginal from the c -bounded average $W^{(c)}(\rho)$ given that the system was initially in state $|s\rangle$. In order to find the fluctuations associated with the optimal c -bounded work extraction protocol we must partition the energy levels into three disjoint subsets $\{1, 2, \dots, d\} = \mathcal{X}_u \cup \mathcal{X}_+ \cup \mathcal{X}_-$, representing the energy levels with positive \mathcal{X}_+ , negative \mathcal{X}_- and unbounded \mathcal{X}_u fluctuations. We also define

$$X_u = \sum_{s \in \mathcal{X}_u} x_s, \quad (11)$$

$$X_{\pm} = \sum_{s \in \mathcal{X}_{\pm}} x_s. \quad (12)$$

The algorithm for determining the partition $\mathcal{X}_u \cup \mathcal{X}_+ \cup \mathcal{X}_-$, which requires the checking of at most $d - 1$ inequalities, is described Appendix 3B. Once we have determined the partition, the fluctuations are given by

$$\theta_s^{(c)} = \begin{cases} \frac{1}{\beta} \log(x_s e^{\beta \mathcal{E}_s}) - \nu, & s \in \mathcal{X}_u \\ +c, & s \in \mathcal{X}_+ \\ -c, & s \in \mathcal{X}_- \end{cases} \quad (13)$$

where

$$\nu = \frac{1}{X_u} (F_u + c(X_+ - X_-)) \quad (14)$$

$$F_u(\rho) = \sum_{s \in \mathcal{X}_u} x_s \log(x_s e^{\beta \mathcal{E}_s}) \quad (15)$$

where $F_u(\rho)$ is the un-normalised free energy calculated for the unbounded partition only. Note that $W^{(c)}(\rho)$ can be written in the more compact form

$$W^{(c)}(\rho) = \frac{1}{\beta} \log \mathcal{Z}' - \frac{1}{\beta} \log (X_u e^{-\beta\nu} + \mathcal{Z}_+ e^{\beta c} + \mathcal{Z}_- e^{-\beta c}) \quad (16)$$

where

$$\mathcal{Z}_{\pm} = \sum_{s \in \mathcal{X}_{\pm}} e^{-\beta \mathcal{E}_s} \quad (17)$$

are the partition functions calculated over the positive and negative bounded partitions respectively. In Appendices 3 and 4 we find that in the optimal work extraction and state formation protocols the final / initial state of the system is the Gibbs state, which is diagonal in the energy eigenbasis. Hence equations (9) and (10) give the optimal work for arbitrary quantum states (See Appendix 1C). For a two-level quantum system $d = 2$, the work content can be expressed succinctly as

$$W^{(c)}(\rho) = \begin{cases} \frac{1}{\beta} \log \mathcal{Z}' - \frac{1}{\beta} \log [e^{+\beta c} (1 + e^{-\beta(\mathcal{E}+c/x_1)})] & \text{if } c < \xi \\ \frac{1}{\beta} \log \mathcal{Z}' - \frac{1}{\beta} \log [e^{-\beta c} (1 + e^{-\beta(\mathcal{E}-c/x_1)})] & \text{if } c > -\xi \\ \frac{1}{\beta} \log \mathcal{Z}' + F(\rho) & \text{otherwise} \end{cases} \quad (18)$$

where

$$\xi = \frac{1}{\beta} \ln(1-x_1) - F(\rho). \quad (19)$$

Without loss of generality we have assumed above $x_1 \geq x_2$ and we define $\mathcal{E} = \mathcal{E}_1 - \mathcal{E}_2$. The above expression derives from (16) and the partitioning algorithm given in Appendix 3.

Returning to the c -bounded work of formation, (10) and the algorithm for finding the state space partition giving the c -bounded work of formation (10) is detailed in Appendix 4. Note that the hamiltonian of ρ , the final state of the system, is given by $H'_S = \sum_j \mathcal{E}'_j |j\rangle\langle j|$. The work of formation for a two level system can be succinctly stated as

$$W_F^{(c)}(\rho) = \begin{cases} \frac{1}{\beta} \log \mathcal{Z} - F(\rho), & c \geq \xi \\ \frac{1}{\beta} \log \left(x e^{\beta \mathcal{E}'} \mathcal{Z} \right) + c \frac{1-x}{x}, & c < \xi \end{cases} \quad (20)$$

where $\rho = x |0\rangle\langle 0| + (1-x) |1\rangle\langle 1|$, $x \geq 1/2$ and $H_S = \mathcal{E} |0\rangle\langle 0|$ is the Hamiltonian of the initial Gibbs state.

We now summarise some of the properties of the c -bounded work.

Result 3. *The c -bounded work is related to the non-fluctuating work by inequalities*

$$W^{(c)}(\rho) \leq W^{(0)}(\rho) + c \quad (21)$$

$$W_F^{(c)}(\rho) \geq W_F^{(0)}(\rho) - c \quad (22)$$

becoming strict inequalities for $c > 0$

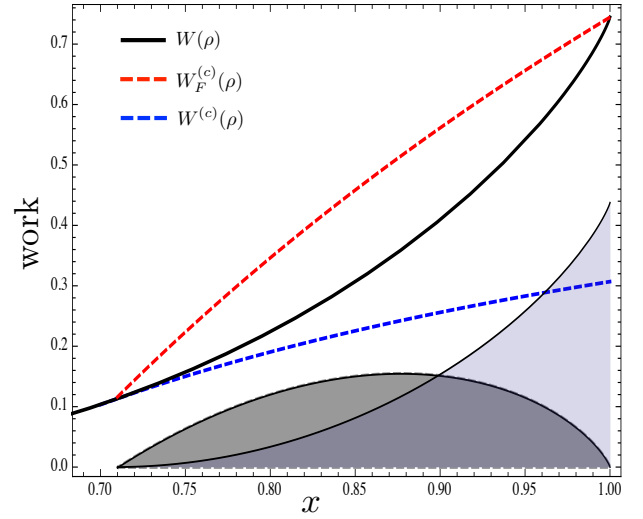


FIG. 1: figure shows the unbounded work $W(\rho)$, the c -bounded work content $W^{(c)}(\rho)$ and the c -bounded work of formation $W_F^{(c)}(\rho)$ for the state $\rho = x |0\rangle\langle 0| + (1-x) |1\rangle\langle 1|$ with Hamiltonian $H_s = \mathcal{E} |0\rangle\langle 0|$ with $\beta = 1$, $\mathcal{E} = 0.1$ and $c = 0.7$. The red shaded area gives the average work dissipated forming the state due to bounding fluctuations, $W_F^{(c)}(\rho) - W^{(\infty)}(\rho)$, and the blue shaded area gives the average work dissipated when extracting c -bounded work, $W^{(c)}(\rho) - W^{(\infty)}(\rho)$. Note that despite choosing a c -bound that allows for fluctuations of the order of the maximal work that can be extracted from the pure state $x = 1$, in general we can extract much less than this amount. There is a discontinuity in $W^{(c)}(\rho)$ at $x = 0$ where we recover $W^{(c)}(\rho) = W^{(\infty)}(\rho)$. Notice also that closer to the thermal state the dissipation is greater for state formations than work extraction, and this reverses as the state becomes more athermal.

Proof. See Appendix 5B □

These inequalities imply that in order to do better than single shot work extraction / state formation, our work must have fluctuations that are greater than the increase in work / decrease in work cost with respect to the deterministic work. In Appendix 7 we show that the c -bounded work distributions that give $W^{(c)}(\rho)$ and $W_F^{(c)}(\rho)$ obey the Jarzynski equality [14]. In Appendix 5 we prove that $\lim_{c \rightarrow 0} W^{(c)}(\rho) = W^{(0)}(\rho)$ and similarly for $W_F^{(c)}(\rho)$.

For the interested reader, it is simple to see that for any finite c a partial ordering of the states w.r.t work emerges. A simple way to observe this is to choose a qubit state ρ with Hamiltonian H_S and a thermal qubit state γ with Hamiltonian $H'_S \neq H_S$ such that neither state thermo-majorizes the other (see [1] for examples). For any two such states there is a value of c below which $W^{(c)}(\rho)$ (for $\rho \rightarrow \gamma$) is negative and $W_F^{(c)}(\rho)$ (for $\gamma \rightarrow \rho$) is positive, i.e. it costs work to perform both the forward and backwards transitions. Note that there are states with the same standard free energy that exhibit a partial order for finite c . Allowing the weight to fluctuate allows us to transition between these states freely. This is an example of how weight is not just a resource for extracting additional fluctuating work but in accommodating dynamics, even when on average its

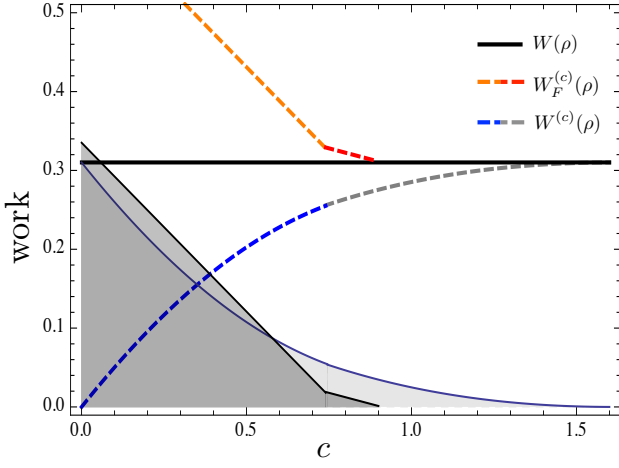


FIG. 2: figure shows $W_F^{(c)}(\rho)$ and $W^{(c)}(\rho)$ v.s. c for the trit state $\rho = 0.7|0\rangle\langle 0| + 0.2|1\rangle\langle 1| + 0.1|2\rangle\langle 2|$ with Hamiltonian $H_s = \mathcal{E}_1|0\rangle\langle 0| + \mathcal{E}_2|1\rangle\langle 1|$ with $\mathcal{E}_1 = 0.1$, $\mathcal{E}_2 = 0.2$. β is set to 1. The red shaded area gives the work dissipated in forming the state, $W_F^{(c)}(\rho) - W^{(\infty)}(\rho)$, and the blue shaded area gives the work dissipated when extracting work from the state, $W(\rho) - W^{(c)}(\rho)$. The different colours represent different partitions used in calculating the bounded work. For small c the formation protocol is more dissipative than the extraction protocol, and for large c the relationship is inverted. Note that for $c > 0.9$ it is possible to reversibly prepare state ρ but not to reversibly extract work from it.

displacement remains zero. It is an interesting open question to determine how much we must allow work to fluctuate to allow a transition $\rho \rightarrow \rho'$ to be achieved without costing work.

Qubit Carnot engine. In this section we find the c -bounded Carnot efficiency for a qubit Carnot engine model, i.e. the maximal efficiency the qubit engine can reach given that fluctuations in the work it produces are bounded by c . We use the same single qubit engine model as described in [2]. The engine operates by moving a qubit ρ with Hamiltonian $H_s = \mathcal{E}|0\rangle\langle 0|$ between two baths of inverse temperature β_H and β_C , with $\beta_H < \beta_C$. The qubit has state $\rho_{H,C} = \mathcal{Z}_{H,C}^{-1} e^{-\beta_H \mathcal{E}} |0\rangle\langle 0| + \mathcal{Z}_{H,C}^{-1} |1\rangle\langle 1|$ when in equilibrium with the hot / cold bath, where $\mathcal{Z}_{H,C} = 1 + e^{-\beta_H \mathcal{E}}$. The engine cycle begins with the qubit in thermal equilibrium with the cold bath. In the first half of the cycle it is then placed in contact with the hot bath and work is extracted. In the second step of the cycle the qubit is returned to the cold bath and work is extracted a second time. In Appendix 6 we show that, in the case that fluctuations are not bounded, it is possible to reach Carnot efficiency with this engine, as shown in [2].

$$\eta_{\text{Carnot}} = 1 - \frac{\beta_H}{\beta_C} \quad (23)$$

In the case that c is finite, the work extracted in the first

half of the cycle is given by

$$W_1^{(c)}(\rho) = \begin{cases} \frac{1}{\beta_H} \log \left(\frac{\mathcal{Z}_H}{\mathcal{Z}_C} \right) + \text{tr}[H_s \rho_C] \left(\frac{\beta_H - \beta_C}{\beta_H} \right), & \text{if } A > c \\ \frac{1}{\beta_H} \log \left(\frac{\mathcal{Z}_H e^{\beta_H c}}{e^{c\beta_H \mathcal{Z}_C} + e^{-\mathcal{E}\beta_H}} \right), & \text{if } A \leq c \end{cases} \quad (24)$$

where

$$A = \frac{\mathcal{E}}{\mathcal{Z}_C} \left(\frac{\beta_C - \beta_H}{\beta_H} \right) \quad (25)$$

if $A \leq c$ we simply extract the difference in free energy between the two thermal states, otherwise we extract the c -bounded work of ρ_C in contact with the bath β_H . Similarly, on the second part of the cycle we extract

$$W_2^{(c)}(\rho) = \begin{cases} \frac{1}{\beta_C} \log \left(\frac{\mathcal{Z}_C}{\mathcal{Z}_H} \right) + \text{tr}[H_s \rho_H] \left(\frac{\beta_C - \beta_H}{\beta_H} \right), & \text{if } B > c \\ \frac{1}{\beta_C} \log \left(\frac{\mathcal{Z}_C e^{-\beta_C c}}{e^{-c\beta_C \mathcal{Z}_H} + e^{-\mathcal{E}\beta_C}} \right), & \text{if } B \leq c \end{cases} \quad (26)$$

where

$$B = \frac{\mathcal{E}}{\mathcal{Z}_H} \left(\frac{\beta_C - \beta_H}{\beta_C} \right) \quad (27)$$

Note that satisfying (27) implies that (25) is also satisfied, therefore breaking inequality (25) is the condition for achieving Carnot efficiency in this model. Also note that B gives the minimum worst case fluctuation of the work extracted by this engine when operating thermodynamically reversibly. The efficiency is given by the ratio of the heat flow from the hot bath to the total work extracted in the cycle. The heat flow from the hot bath is found by applying the 1st law of thermodynamics, $Q_H = \Delta\langle \mathcal{E} \rangle(\rho_H \rightarrow \rho_C) + W_1^{(c)}$ where $\Delta\langle \mathcal{E} \rangle(\rho_H \rightarrow \rho_C)$ is the change in the systems internal energy in the first part of the cycle. Therefore the c -bounded efficiency of the engine is given by

$$\eta^{(c)} = \frac{W_1^{(c)} + W_2^{(c)}}{\Delta U + W_1^{(c)}} \quad (28)$$

In the case that $c \rightarrow \infty$, we recover the Carnot efficiency, which is bounded from above by 1, i.e. we recover unit efficiency in the limit that $\beta_C/\beta_H \rightarrow \infty$. For any finite c this is no longer the case, with the maximal efficiency given by

$$\eta_{\text{max}}^{(c)} = \lim_{\beta_C/\beta_H \rightarrow \infty} \eta^{(c)} = 1 - \frac{\mathcal{E}}{2(2c + \mathcal{E})} \quad (29)$$

This gives an upper limit on the efficiency of the single qubit engine protocol described above, which is dependent only on the Hamiltonian of the qubit and the parameter c .

As $\mathcal{E} \rightarrow 0$, $\eta_{\text{max}}^{(c)} \rightarrow 1$, but the work extracted tends to zero as the Gibbs states associated with the two bath temperatures become indistinguishable. For $c \rightarrow 0$ we get that $\eta_{\text{max}}^{(c)}$ is bounded from below by 1/2, but at $c = 0$ no work can be extracted as the thermal states are of full rank, giving $\eta_{\text{max}}^{(0)} = 0$. Therefore we find that, although this engine cannot run at non-zero efficiency in the single-shot regime, if we allow for arbitrarily

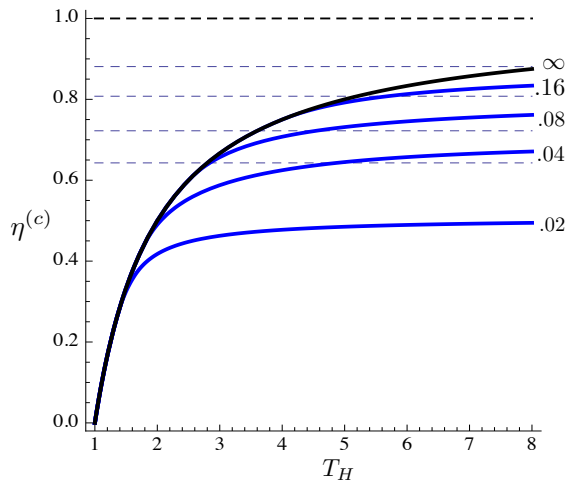


FIG. 3: Figure shows the c -bounded Carnot efficiencies vs. T_H for single qubit engine with gap $\mathcal{E} = 0.1$ and $T_C = 1$. The black line gives the unbounded Carnot efficiency. The Blue lines give the c -bounded efficiencies with there corresponding c 's marked on the figure. The Black dashed lines give the maximum attainable efficiency in the limit of asymptotic temperature difference between $\beta_C/\beta_H \rightarrow \infty$

small fluctuations it is possible in principle to reach a maximum efficiency greater than $1/2$ (although the fluctuations in work will still be of the order of the work extracted, given inequality (21)). Similar results relating to the single-shot regime are discussed in [15].

Conclusions. In this article we have derived tight bounds on the minimal fluctuations in work associated with thermodynamically reversible protocols, for which the average work is given by difference in free energy between initial and final states. We have found that thermodynamically reversible protocols have fluctuations that diverge in size as the relative athermality of initial or final states increases.

Motivated by this we have presented a powerful framework for computing the work associated with a thermodynamic process under arbitrary convex bounds. We have derived the c -bounded work content and work of formation of arbitrary quantum states, which can be understood as modified free energies that interpolate between the standard and single shot free energies. By exploring this new territory, we have found that the phenomenology of single-shot thermodynamics, namely thermodynamic irreversibility and a partial order of states with respect to work, are to some extent present for any finite c . Furthermore we have found that it is

impossible to extract more than the deterministic work content of a system without necessitating fluctuations that are greater than the gain in work (and similarly for the work cost of state formation).

An interesting open question is to what extent we must allow work to fluctuate in order to allow for a given state transformation. Answering this question would require the extension of the results presented in this article to processes with arbitrary initial and final states, including the case where both initial and final states contain coherences between energy levels. In Appendix 1C we show that, under the assumption that the protocol is independent on the position state of the weight, the “coherence modes” [16–18] evolve independently under the action of the thermal map. This lays the ground for future investigations into how the presence of coherences affects the allowed thermal operations in the case that work is allowed to fluctuate.

Finally, we have used the c -bounded work to study how bounding work fluctuations affects the efficiency of a single qubit nano-engine, and have derived an upper bound on efficiency of this engine that depends only on c and the engine’s Hamiltonian, establishing a fundamental trade-off between the engine’s efficiency and the fluctuations in the work it produces. This opens the door to correcting the efficiency for general thermodynamic protocols, taking into account the fragility of realistic machines that cannot tolerate large fluctuations in work.

Given that there are many thermal engine models that can reach Carnot efficiency in the case the fluctuations in work are unbounded, it would be of interest to determine the optimal engine with respect to minimizing fluctuations in work whilst maximising efficiency or the power produced. Furthermore, it is well known that in the thermodynamic limit the relative size of fluctuations in work to the average tends to zero. It would be of interest to determine if it is possible to design engines operating far from the thermodynamic limit that achieve a similar quasi-deterministic work output with non-zero power. For example it could be possible, through clever choice of the working system Hamiltonian, or by controlling interactions between a small number of systems that constitute the working system, to find engine models that achieve quasi-deterministic work output without needing to take the thermodynamic limit. Further work in this direction would provide invaluable insights for designing realistic nano-engines that are robust to fluctuations.

[1] Michał Horodecki and Jonathan Oppenheim. Fundamental limitations for quantum and nanoscale thermodynamics. *Nature communications*, 4, 2013.
[2] Paul Skrzypczyk, Anthony J Short, and Sandu Popescu. Work extraction and thermodynamics for individual quantum systems. *Nature communications*, 5, 2014.
[3] J Gemmer and J Anders. From single-shot towards general work extraction in a quantum thermodynamic framework. *arXiv preprint arXiv:1504.05061*, 2015.

[4] Johan Åberg. Truly work-like work extraction via a single-shot analysis. *Nature communications*, 4, 2013.
[5] Oscar CO Dahlsten, Renato Renner, Elisabeth Rieper, and Vlatko Vedral. Inadequacy of von neumann entropy for characterizing extractable work. *New Journal of Physics*, 13(5):053015, 2011.
[6] Noah Linden, Sandu Popescu, and Paul Skrzypczyk. How small can thermal machines be? the smallest possible refrigerator. *Physical review letters*, 105(13):130401, 2010.

- [7] CA Ryan, O Moussa, J Baugh, and R Laflamme. Spin based heat engine: demonstration of multiple rounds of algorithmic cooling. *Physical review letters*, 100(14):140501, 2008.
- [8] HT Quan, YD Wang, Yu-xi Liu, CP Sun, and Franco Nori. Maxwells demon assisted thermodynamic cycle in superconducting quantum circuits. *Physical review letters*, 97(18):180402, 2006.
- [9] Fernando Brandão, Michał Horodecki, Nelly Ng, Jonathan Oppenheim, and Stephanie Wehner. The second laws of quantum thermodynamics. *Proceedings of the National Academy of Sciences*, 112(11):3275–3279, 2015.
- [10] Lluís Masanes and Jonathan Oppenheim. A derivation (and quantification) of the third law of thermodynamics. *arXiv preprint arXiv:1412.3828*, 2014.
- [11] R Gallego, J Eisert, and H Wilming. Defining work from operational principles. *arXiv preprint arXiv:1504.05056*, 2015.
- [12] Yūto Murashita and Masahito Ueda. Absolute irreversibility resolves the gibbs paradox. *arXiv preprint arXiv:1506.04468*, 2015.
- [13] Yūto Murashita, Ken Funo, and Masahito Ueda. Nonequilibrium equalities in absolutely irreversible processes. *Physical Review E*, 90(4):042110, 2014.
- [14] Christopher Jarzynski. Nonequilibrium equality for free energy differences. *Physical Review Letters*, 78(14):2690, 1997.
- [15] Mischa P Woods, Nelly Ng, and Stephanie Wehner. The maximum efficiency of nano heat engines depends on more than temperature. *arXiv preprint arXiv:1506.02322*, 2015.
- [16] Matteo Lostaglio, David Jennings, and Terry Rudolph. Description of quantum coherence in thermodynamic processes requires constraints beyond free energy. *Nature communications*, 6, 2015.
- [17] Matteo Lostaglio, Kamil Korzekwa, David Jennings, and Terry Rudolph. Quantum coherence, time-translation symmetry, and thermodynamics. *Physical Review X*, 5(2):021001, 2015.
- [18] Piotr Ćwikliński, Michał Studziński, Michał Horodecki, and Jonathan Oppenheim. Limitations on the evolution of quantum coherences: Towards fully quantum second laws of thermodynamics. *Physical review letters*, 115(21):210403, 2015.
- [19] Johan Aberg. Fully quantum fluctuation theorems. *arXiv preprint arXiv:1601.01302*, 2016.
- [20] Álvaro M Alhambra, Lluís Masanes, Jonathan Oppenheim, and Christopher Perry. The second law of quantum thermodynamics as an equality. *arXiv preprint arXiv:1601.05799*, 2016.

I. PRELIMINARIES

A. Thermal operations with fluctuating work

In this appendix we introduce the framework that we will use to derive $W^{(c)}(\rho)$. The work extraction protocol is performed using a system ρ , a infinite thermal bath and a work system or *weight*.

First, let us characterize the type of process/operation that we consider, which we refer to as *thermal operations with fluctuating work*. Our setting consists of a system with Hamiltonian H_S , a bath with Hamiltonian H_B initially in the thermal state $\rho_B = \frac{1}{Z_B} e^{-\beta H_B}$, and an ideal

weight with Hamiltonian $H_W = \int_{\mathbb{R}} dx x |x\rangle\langle x|$, where the orthonormal basis $\{|x\rangle, \forall x \in \mathbb{R}\}$ represents the position of the weight. Any joint transformation of system, bath and weight is represented by a Completely Positive Trace Preserving (CPTP) map Γ_{SBW} satisfying the following conditions:

Microscopic reversibility (Second Law): It has an (CPTP) inverse Γ_{SBW}^{-1} , which implies unitarity $\Gamma_{\text{SBW}}(\rho_{\text{SBW}}) = U \rho_{\text{SBW}} U^\dagger$.

Energy conservation (First Law): $[U, H_S + H_B + H_W] = 0$.

Independence from the “position” of the weight: The unitary commutes with the translations on the weight $[U, \Delta_W] = 0$. The generator of the translations Δ_W is canonically conjugated to the position of the weight $[H_W, \Delta_W] = i$.

Classicality of work: Before and after applying the global map Γ_{SBW} the position of the weight is measured, obtaining outcomes $|x\rangle$ and $|x+w\rangle$ respectively. The joint transformation of the system and work random variable w is given by the map

$$\Lambda(\rho_S, w) = \int_{\mathbb{R}} dx \text{tr}_{\text{BW}} [Q_{x+w} U (\rho_S \rho_B Q_x \rho_W Q_x) U^\dagger] , \quad (30)$$

where $Q_x = |x\rangle\langle x|$ is a weight eigen-projector.

Condition $[U, \Delta_W] = 0$ implies that the reduced map on system and bath is a mixture of unitaries (Result 1 in [10]). Hence, this transformation can never decrease the entropy of system and bath, which guarantees that the weight is not used as a source of non-equilibrium. Note that after integrating over the work variable

$$\int_{\mathbb{R}} dw \Lambda(\rho_S, w) = \text{tr}_{\text{BW}} [U (\rho_S \rho_B \Theta[\rho_W]) U^\dagger] = \Gamma_S(\rho_S) , \quad (31)$$

we obtain the reduced CPTP map for the system. The state $\Theta[\rho_W] = \int_{\mathbb{R}} dx Q_x \rho_W Q_x$ is the energy-diagonal version of ρ_W . But $\Lambda(\rho_S, w)$ being independent of ρ_W , we can choose the initial state to be diagonal $\Theta[\rho_W]$. Also, note that when tracing the system

$$\text{tr}_S \Lambda(\rho_S, w) = P(w) , \quad (32)$$

we obtain the probability distribution of the work generated in the transformation.

B. Thermal operations with non-constant Hamiltonian

Thermal operations are general enough to include the case where the initial Hamiltonian of the system H_S is different than the final one H'_S . This is done by including an additional qubit X which plays the role of a switch (as in [1, 19]). Now the total Hamiltonian is

$$H = H_S \otimes |0\rangle_X\langle 0| + H'_S \otimes |1\rangle_X\langle 1| + H_B + H_W , \quad (33)$$

and energy conservation reads $[V, H] = 0$, where V is the global unitary when we include the switch. We impose that the initial state of switch is $|0\rangle_X$ and the global unitary V performs the switching

$$V(\rho_{\text{SBW}} \otimes |0\rangle_X \langle 0|) V^\dagger = \rho'_{\text{SBW}} \otimes |1\rangle_X \langle 1|, \quad (34)$$

for any ρ_{SBW} . This implies

$$V = U \otimes |1\rangle_X \langle 0| + \tilde{U} \otimes |0\rangle_X \langle 1|, \quad (35)$$

where U and \tilde{U} are unitaries on system, bath and weight. Condition $[V, H] = 0$ implies

$$U(H_S + H_B + H_W) = (H'_S + H_B + H_W)U. \quad (36)$$

Therefore, the reduced map on system, bath and weight can be written as

$$\Gamma_{\text{SBW}}(\rho_{\text{SBW}}) = U \rho_{\text{SBW}} U^\dagger, \quad (37)$$

where the unitary U does not necessarily commute with $H_S + H_B + H_W$ nor $H'_S + H_B + H_W$ but satisfies (36).

C. Reducing the quantum problem to the classical case

Let us show that the equation connecting the initial state of the system ρ_S with the final one conditioned on work w ,

$$\rho'_{S|w} = \frac{1}{P(w)} \Lambda(\rho_S, w), \quad (38)$$

decouples in the diagonal part and the other energy modes [16–18]. The map Θ_α defined as

$$\Theta_\alpha[\rho_S] = \int_{\mathbb{R}} dt e^{i\alpha t} e^{iH_S t} \rho_S e^{-iH_S t}, \quad (39)$$

projects the state ρ_S onto the α -energy mode of H_S . When $\alpha = 0$ it projects the state onto its diagonal, when written in the energy eigenbasis. And in general, it projects the state onto all the terms $|s_1\rangle\langle s_2|$ such that $\mathcal{E}_{s_1} - \mathcal{E}_{s_2} = \alpha$. Using constraint (36) and identities $e^{itH_W} Q_x = e^{itx}$ and $[H_B, \rho_B] = 0$ we obtain

$$\begin{aligned} \Theta_\alpha[\Lambda(\rho_S, w)] &= \int_{\mathbb{R}} dt e^{i\alpha t} e^{iH'_S t} \Lambda(\rho_S, w) e^{-iH'_S t} \\ &= \int_{\mathbb{R}} dt dx e^{i\alpha t} \text{tr}_{\text{BW}} \left[e^{i(H'_S + H_B + H_W)t} Q_{x+w} U (\rho_S \rho_B Q_x \rho_W Q_x) U^\dagger e^{-i(H'_S + H_B + H_W)t} \right] \\ &= \int_{\mathbb{R}} dt dx e^{i\alpha t} \text{tr}_{\text{BW}} \left[Q_{x+w} U \left(e^{i(H_S + H_B + H_W)t} \rho_S \rho_B Q_x \rho_W Q_x e^{-i(H_S + H_B + H_W)t} \right) U^\dagger \right] \\ &= \int_{\mathbb{R}} dt dx e^{i\alpha t} \text{tr}_{\text{BW}} \left[Q_{x+w} U \left(e^{iH_S t} \rho_S e^{-iH_S t} \rho_B Q_x \rho_W Q_x \right) U^\dagger \right] \\ &= \Gamma_S(\Theta_\alpha[\rho_S], w), \end{aligned} \quad (40)$$

as claimed above. This shows that when the initial state is diagonal, $\Theta_\alpha[\rho_S] = 0$ for all $\alpha \neq 0$, so is the final one; and visa-versa. Thus our results, concerning work extraction and state formation where either the initial or final state diagonal in the energy eigenbasis, are valid in the case that the non-equilibrium state involves coherences between energy eigenstates.

The diagonal part of the map is nicely characterized by the conditional probability distribution

$$t(s', w | s) = \langle s' | \Lambda(|s\rangle\langle s|, w) | s' \rangle, \quad (41)$$

where $|s\rangle$ is the eigenbasis of $\Theta[\rho_S]$ and $|s'\rangle$ is the eigenbasis of $\Theta[\Lambda(\rho_S, w)]$. Note that the “dynamics” of the diagonals is a completely classical problem. From now on we use $\rho = \Theta[\rho_S]$, $\rho' = \Gamma_S \Theta[\rho_S]$ for the initial/final states in optimal work extraction / state formation processes.

D. Necessary condition for thermal operations

Using (36) and definitions made above we obtain

$$\begin{aligned}
& \sum_s \int_{\mathbb{R}} dw t(s', w|s) e^{\beta(\mathcal{E}_{s'} - \mathcal{E}_s + w)} \\
&= \sum_s \int_{\mathbb{R}} dw dx \langle s' | e^{\beta H'_S} \text{tr}_{\text{BW}} [e^{\beta H_W} Q_{x+w} U (e^{-\beta H_S} |s\rangle\langle s| \rho_B e^{-\beta H_W} Q_x \rho_W Q_x) U^\dagger] |s'\rangle \\
&= \int_{\mathbb{R}} dx' dx \langle s' | e^{\beta H'_S} \text{tr}_{\text{BW}} \left[e^{\beta H_W} Q_{x'} U \left(e^{-\beta H_S} \frac{e^{-\beta H_B}}{Z_B} e^{-\beta H_W} Q_x \rho_W Q_x \right) U^\dagger \right] |s'\rangle \\
&= \langle s' | e^{\beta H'_S} \text{tr}_{\text{BW}} \left[e^{\beta H_W} e^{-\beta H'_S} \frac{e^{-\beta H_B}}{Z_B} e^{-\beta H_W} U \Theta[\rho_W] U^\dagger \right] |s'\rangle \\
&= \text{tr}_{\text{SBW}} \left[|s'\rangle\langle s'| \frac{e^{-\beta H_B}}{Z_B} U \Theta[\rho_W] U^\dagger \right].
\end{aligned} \tag{42}$$

As mentioned above, Result 1 in [third law masanes oppenheim] proves that condition $[U, \Delta_W] = 0$ implies

$$\text{tr}_W [U \mathbb{I}_{\text{SB}} \rho_W U^\dagger] = \mathbb{I}_{\text{SB}} \tag{43}$$

for all ρ_W . Applying this to (42) we obtain

$$\text{tr}_{\text{SBW}} \left[|s'\rangle\langle s'| \frac{e^{-\beta H_B}}{Z_B} U \Theta[\rho_W] U^\dagger \right] = \text{tr}_{\text{SB}} \left[|s'\rangle\langle s'| \frac{e^{-\beta H_B}}{Z_B} \right] = 1. \tag{44}$$

This proves the “only if” part of the following

Result 4. *The (classical) map $t(s', w|s)$ comes from a thermal operation*

$$t(s', w|s) = \langle s' | \Lambda(|s\rangle\langle s|, w) |s'\rangle, \tag{45}$$

if and only if

$$\sum_s \int_{\mathbb{R}} dw t(s', w|s) e^{\beta(\mathcal{E}_{s'} - \mathcal{E}_s + w)} = 1, \tag{46}$$

for all s' .

The “if” part of the above result is proven in Result 5 from [20].

E. Optimal thermal operations have minimal work fluctuations

Given a process $t(s', w|s)$, satisfying (46), the work generated in a particular state transition $s \rightarrow s'$ has probability distribution

$$t(w|s, s') = \frac{t(s', w|s)}{t(s'|s)}, \tag{47}$$

where $t(s'|s) = \int_{\mathbb{R}} dw t(s', w|s)$ is the marginal transformation on the system. In general, despite the conditioning on s, s' , the distribution (47) contains fluctuations on w . In certain setups, a less fluctuating work variable w is desirable. The following result shows that this can always be done without decreasing the average work generated in the given process $t(s', w|s)$.

Result 5. *If $t(s', w|s)$ satisfies condition (46) then*

$$\tilde{t}(s', w|s) = \delta(w - w_{s,s'}) t(s'|s) \tag{48}$$

with

$$w_{s,s'} = T \ln \int_{\mathbb{R}} dw t(w|s, s') e^{\beta w} \tag{49}$$

also satisfies (46), and in addition

$$\langle w \rangle_{\tilde{t}} \geq \langle w \rangle_t, \tag{50}$$

$$\max_{w: \tilde{t}(w) > 0} |w - \langle w \rangle_{\tilde{t}}| \leq \max_{w: t(w) > 0} |w - \langle w \rangle_t|. \tag{51}$$

The above inequalities are saturated if and only if $\tilde{t}(w, s'|s) = t(w, s'|s)$.

To show that $\tilde{t}(s', w|s)$ satisfies (46), first, exponentiate the two sides of (49),

$$e^{\beta w_{s,s'}} = \int_{\mathbb{R}} dw t(w|s, s') e^{\beta w}, \quad (52)$$

and second, multiply by $t(s'|s) e^{\beta(\mathcal{E}_{s'} - \mathcal{E}_s)}$ and sum over s ,

$$\sum_s t(s'|s) e^{\beta(\mathcal{E}_{s'} - \mathcal{E}_s + w_{s,s'})} = \sum_s \int_{\mathbb{R}} dw t(s', w|s) e^{\beta(\mathcal{E}_{s'} - \mathcal{E}_s + w)} = 1. \quad (53)$$

Note that the two maps, $t(s', w|s)$ and $\tilde{t}(s', w|s)$, have the same marginal $t(s'|s)$. That is, they perform the same transformation on the system.

To show (50) we use the convexity of the exponential in equation (49), obtaining

$$w_{s,s'} \geq T \ln \int_{\mathbb{R}} dw P(w|s, s') e^{\beta w} = \int_{\mathbb{R}} dw t(w|s, s') w. \quad (54)$$

Averaging over s, s' gives (50). Also, note that due to strict convexity of the exponential, the equality in (54) is only achieved when $t(w|s, s') = \delta(w - w_{s,s'}) = \tilde{t}(w|s, s')$.

To show (51), note that the convexity of the exponential implies that, unless $t(w|s, s') = \delta(w - w_{s,s'})$, there are values $w_+ > w_{s,s'}$ and $w_- < w_{s,s'}$ such that $t(w_{\pm}|s, s') > 0$. Hence, equality is only achieved when $t = \tilde{t}$.

II. OPTIMAL UNBOUNDED WORK

In this appendix we calculate the maximal average work extracted (or minimal work of formation) for the state transformation $\rho \rightarrow \rho'$ with unbounded fluctuations. The work is given by the difference in free energy. By explicitly calculating the work distribution we show that the work fluctuations diverge as the initial or final system states become more athermal. We also show that there is no map that can achieve the optimal average work (i.e. thermodynamically reversible) with a smaller range of fluctuations about the average (see also Appendix 1E).

In the previous Appendices we simplified the work-optimal map to a form where it is defined by the map parameters $\{t(s'|s), w_{ss'}\}$ where $t(s'|s)$ defined the reduced map on the system. We have shown that, in the case that the initial or final state is diagonal we can work with de-phased initial and final states. Furthermore the $t(s'|s)$ must obey

$$1 = \sum_{s'} t(s'|s) \quad (55)$$

$$x_{s'} = \sum_s x_s t(s'|s) \quad (56)$$

$$1 = \sum_s t(s'|s) e^{\beta(\mathcal{E}_{s'} - \mathcal{E}_s + w_{s,s'})} \quad (57)$$

where the first two constraints ensure that the reduced map acting on the system is stochastic and achieves the desired state transformation, and the third constraint, derived in the previous section, is required to ensure that the map is microscopically reversible. In the following we simplify the map further in the case of optimal work extraction.

It will be useful to relax the reversibility equality to an inequality

$$1 \geq \sum_s t(s'|s) e^{\beta(\mathcal{E}_{s'} - \mathcal{E}_s + w_{s,s'})} \quad (58)$$

where saturation of the inequality implies that the map is an allowed (thermal) operation. In all future calculations we make use of this relaxed constraint and show that our solutions saturate the inequality.

The average work associated with the optimal map achieving $\rho \rightarrow \rho'$ is given by

$$W = \sum_{ss'} x_s t(s'|s) w_{ss'} \quad (59)$$

it turns out to be sufficient to optimize this under the reversibility constraint (58). The Lagrangian is

$$\mathcal{L} = \sum_{ss'} x_s t(s'|s) w_{ss'} + \sum_{s'} \lambda_{s'} \left(1 - \sum_{ss'} t(s'|s) e^{\beta(w_{ss'} + \Delta \mathcal{E}_{ss'})} \right) \quad (60)$$

Maximising with respect to $w_{ss'}$ gives

$$w_{ss'} = \frac{1}{\beta} \log \left(\frac{x_s e^{\beta \Delta \mathcal{E}_{ss'}}}{\lambda_{s'} \beta} \right) \quad (61)$$

Extremizing with respect to $\lambda_{s'}$ and applying (55) and (56) gives

$$\lambda_{s'} = \frac{x_{s'}}{\beta} \quad (62)$$

Substituting this into (61) gives the optimal work

$$\sum_{ss'} x_s t(s'|s) \frac{1}{\beta} \log \left(\frac{x_s e^{\beta \Delta \mathcal{E}_{ss'}}}{x_{s'}} \right) = F(\rho) - F(\rho') \quad (63)$$

where $F(\rho) = \langle \mathcal{E}(\rho) \rangle - 1/\beta S(\rho)$ is the standard free energy and we have again used (55), (56). Notice that the result is independent of our choice of $t(s'|s)$, i.e. any map that takes us from $\rho \rightarrow \rho'$ gives the same optimal work. This is simply a statement that the free energy is a state variable (i.e. is path independent). The work $W^{(\infty)}$ is the average of the work marginals, gives by

$$w_{ss'} = \frac{1}{\beta} \log \left(\frac{x_s e^{\beta \Delta \mathcal{E}_{ss'}}}{x_{s'}} \right) \quad (64)$$

Note work marginals with $x_s = 0$ are set to zero (they have zero probability of occurring in the work distribution). Note that the fluctuations diverge as the initial / final state moves further from equilibrium. It is easy to check that substituting our solutions for $w_{ss'}$ into (58) saturates the inequality, therefore this map is achievable with thermal operations.

Result 5 shows that no thermal map can achieve this W with smaller worst case fluctuations. Having explicitly calculated the work marginals, we have derived necessary and sufficient conditions for a thermal map to exist that achieves this W (i.e. the thermodynamically reversible work) given that the fluctuations are constrained $|w - W| \leq c$

Result 1. The process $(\rho, H_S) \rightarrow (\rho', H'_S)$ can be achieved in a thermodynamically reversible map if

$$e^{\beta(\Delta F(\rho \rightarrow \rho') - c)} \leq \frac{x_s e^{\beta \mathcal{E}_s}}{x_{s'} e^{\beta \mathcal{E}_{s'}}} \leq e^{\beta(\Delta F(\rho \rightarrow \rho') + c)} \quad \forall s, s' \quad (65)$$

This becomes a necessary and sufficient condition if the initial and/or or final state is diagonal in the energy eigenbasis

Proof. Given Result 5, it is sufficient to find the conditions that the work marginals (64) obey the c -bound

$$\left| \frac{1}{\beta} \log \left(\frac{x_s e^{\beta \Delta \mathcal{E}_{ss'}}}{x_{s'}} \right) - \Delta F(\rho \rightarrow \rho') \right| \leq c \quad (66)$$

which gives the desired inequalities \square

III. OPTIMAL WORK EXTRACTION WITH BOUNDED FLUCTUATIONS

In this appendix we derive the c -bounded work content for general quantum state ρ .

A. determining the optimal map

Lemma 1. *We can always find an optimal protocol where $t(s'|s) \rightarrow \tilde{t}_s = e^{-\beta \mathcal{E}_{s'}} / \mathcal{Z}'$, where $\mathcal{Z}' = \text{tr}[e^{-\beta H'_S}]$, and $w_{ss'} \rightarrow \tilde{w}_s$ that obeys $|\tilde{w}_s - \langle W \rangle| \leq |w_{ss'} - \langle W \rangle| \quad \forall s, s'$*

Proof. The average work extracted by a given protocol is given by

$$W = \sum_{ss'} x_s t(s'|s) w_{ss'} \quad (67)$$

where $\{t(s'|s), w_{ss'}\}$ obey constraints (55) (56) and (58).

Assume there exists some optimal protocol $\{t(s'|s), w_{ss'}\}$ that satisfies these and obeys the c -bound

$$|w_{ss'} - W| \leq c \quad \forall s, s' \quad (68)$$

Define a new protocol with

$$\tilde{t}_{s'} = \frac{e^{-\beta \mathcal{E}_{s'}}}{\mathcal{Z}'} \quad (69)$$

$$\tilde{w}_s = \sum_k t(k|s) w_{sk} \quad (70)$$

The average work extracted by this protocol is

$$\tilde{W} = \sum_{ss'} x_s \tilde{t}_{s'} \tilde{w}_s = \sum_s x_s \tilde{w}_s = \sum_{ss'} x_s t(s'|s) w_{ss'} \quad (71)$$

therefore it extract the same amount of work as the optimal protocol. It also obeys the microscopic reversibility inequality (58) as

$$\begin{aligned} \sum_s \tilde{t}_{s'} e^{\beta(\tilde{w}_s - \Delta \mathcal{E}_{ss'})} &= \frac{1}{\mathcal{Z}'} \sum_s e^{\beta \left(\sum_k t(k|s) w_{sk} - \mathcal{E}_s \right)} \\ &\leq \frac{1}{\mathcal{Z}'} \sum_{sk} t(k|s) e^{\beta w_{sk} - \beta \mathcal{E}_s} \\ &= \frac{1}{\mathcal{Z}'} \sum_{sk} t(k|s) e^{\beta w_{sk} - \beta \mathcal{E}_s + \beta \mathcal{E}_k - \beta \mathcal{E}_k} \\ &\leq \frac{1}{\mathcal{Z}'} \sum_k e^{-\beta \mathcal{E}_k} \\ &= 1 \end{aligned}$$

where we have used the fact that $\sum_k t(k|s) = 1$ and the convexity of the exponential function to get $e^{\beta \tilde{w}_s} \leq \sum_k t(k|s) e^{\beta w_{sk}}$, and in the second to last line have used the reversibility of the original map. Finally, the new protocol also obeys the c -bound as the work values \tilde{w}_s are convex sums of the work marginals $w_{ss'}$ and therefore $\max\{w_{sk}\} \geq \tilde{w}_s \geq \min\{w_{sk}\}$. As W is unchanged then the worst case fluctuations of the new work distribution about the average must be less than or equal to those of the optimal distribution. \square

We now drop the tilde from \tilde{w}_s and $\tilde{t}_{s'}$. It is simple to check that $t_{s'} = 1/\mathcal{Z}' e^{-\beta \mathcal{E}_{s'}}$ satisfies the conditions (55) and (56). The reversibility inequality (58) becomes

$$\mathcal{Z}' \geq \sum_s e^{\beta(w_s - \mathcal{E}_s)} \quad (72)$$

and the average work is given by

$$W = \sum_s x_s w_s \quad (73)$$

We are now in a position to derive $W^{(c)}(\rho)$. The Lagrangian is the same as employed in the previous section except that it includes terms bounding all $|w_s - \sum_k x_k w_k| \leq 0$. Due to the exponential in the reversibility term $\mathcal{Z}' \geq \sum_s e^{\beta(w_s - \mathcal{E}_s)}$ we must linearise these bounds so as to avoid generating a transcendental equation when extremizing the Lagrangian over w_s . We therefore select the bounds

$$w_s - \sum_k x_k w_k \leq c \quad (74)$$

$$\sum_k x_k w_k - w_s \leq c \quad (75)$$

with associate Lagrange parameters $\mu_{ss'}$ and $\bar{\mu}_{ss'}$. As all fluctuations are either positive ($w_s \geq W^{(c)}(\rho)$) or negative ($w_s < W^{(c)}(\rho)$) one of these bounds will always be trivial for each work marginal w_s . Including the reversibility constraint (72) we maximise the average work $W = \sum_s x_s w_s$ by extremizing the Lagrangian

$$\mathcal{L} = \sum_s f_s w_s + \lambda \left(\mathcal{Z}' - \sum_s e^{\beta(w_s - \mathcal{E}_s)} \right) + c \sum_s (\mu_s + \bar{\mu}_s) \quad (76)$$

where

$$f_s = x_s f - (\bar{\mu}_s - \mu_s) \quad (77)$$

where $f = 1 + \sum_k (\bar{\mu}_k - \mu_k)$ and we have replaced the old Lagrange parameters with $\lambda = \mathcal{Z}' \sum_s \lambda_s$ and $\mu_s(\bar{\mu}_s) = \sum_{s'} \mu_{ss'}(\bar{\mu}_{ss'})$. Maximizing w.r.t w_i and λ gives

$$w_s = \frac{1}{\beta} \log \left(\frac{f_s e^{\beta \mathcal{E}_s}}{\beta \lambda} \right) \quad (78)$$

$$\lambda = \frac{1}{\beta \mathcal{Z}'} \quad (79)$$

where we have used $\sum_s f_s = 1$. Substituting these into the Lagrangian simplifies it to

$$\mathcal{L} = \frac{1}{\beta} \sum_s f_s \log (f_s e^{\beta \mathcal{E}_s} \mathcal{Z}') + c \sum_s (\mu_s + \bar{\mu}_s) \quad (80)$$

Maximizing w.r.t μ_j and $\bar{\mu}_j$ under the condition $\mu_j \geq 0$ and $\bar{\mu}_j \geq 0$ gives

$$(81)$$

$$\frac{\partial \mathcal{L}}{\partial \mu_j} = 0 \rightarrow \log (f_j e^{\beta \mathcal{E}_j}) - \sum_s x_s \log (f_s e^{\beta \mathcal{E}_s}) = -\beta c \quad (82)$$

$$\frac{\partial \mathcal{L}}{\partial \bar{\mu}_j} = 0 \rightarrow -\log (f_j e^{\beta \mathcal{E}_j}) + \sum_s x_s \log (f_s e^{\beta \mathcal{E}_s}) = -\beta c$$

where we have used

$$\frac{\partial f_s}{\partial \mu_j} = -x_s + \delta_{sj} \quad (83)$$

$$\frac{\partial f_s}{\partial \bar{\mu}_j} = x_s - \delta_{sj} \quad (84)$$

For $c > 0$, equations (81) and (82) cannot be simultaneously satisfied by any $w_j = 1/\beta \log (f_j e^{\beta \mathcal{E}_j} \mathcal{Z}')$ therefore if $\mu_j > 0$ then $\bar{\mu}_j = 0$ and visa-versa. As $w_i = 1/\beta \log (f_i e^{\beta \mathcal{E}_i} \mathcal{Z}')$ equations (81) and (82) can be written in the form

$$\mu_j \neq 0 \rightarrow w_j - \langle W \rangle = -c \quad (85)$$

$$\bar{\mu}_j \neq 0 \rightarrow w_j - \langle W \rangle = c \quad (86)$$

Therefore fluctuations saturate either a positive or negative bound, or saturate no bounds ($\mu_j = \bar{\mu}_j = 0$). It will therefore be useful to partition the set of energy levels into those for which the resulting fluctuations will saturate a positive bound $i \in \mathcal{X}_+$, a negative bound $j \in \mathcal{X}_-$ or saturate no bounds $u \in \mathcal{X}_u$. For fluctuations $w_i, w_{i'}$

that saturate a positive bound and $w_j, w_{j'}$ that saturate a negative bound we have

$$f_i e^{\beta \mathcal{E}_i} = f_{i'} e^{\beta \mathcal{E}_{i'}} \quad (87)$$

$$f_j e^{\beta \mathcal{E}_j} = f_{j'} e^{\beta \mathcal{E}_{j'}} \quad (88)$$

$$f_j e^{\beta \mathcal{E}_j} = f_{i'} e^{\beta (\mathcal{E}_{i'} - 2c)} \quad (89)$$

$$f_i = x_i f - \bar{\mu}_i \quad (90)$$

$$f_j = x_j f + \mu_j \quad (91)$$

$$f_u = x_u f \quad (92)$$

where (87) comes from summing (81) for $\mu_j, \mu_{j'}$, (88) from summing (82) for $\bar{\mu}_i, \bar{\mu}_{i'}$, (89) comes from subtracting (81) for μ_j and (82) for $\bar{\mu}_i$ and (90)–(92) are just the definition (77) with the conditions $\bar{\mu}_j = 0, \mu_i = 0$ and $\mu_u = \bar{\mu}_u = 0$ applied. (87)–(89) and (92) allow us to simplify (81) and (82) to

$$\frac{\partial \mathcal{L}}{\partial \mu_j} = 0 \rightarrow \log \left(\frac{f}{f_j e^{\beta \mathcal{E}_j}} \right) = \nu + c \quad (93)$$

$$\frac{\partial \mathcal{L}}{\partial \bar{\mu}_i} = 0 \rightarrow \log \left(\frac{f}{f_i e^{\beta \mathcal{E}_i}} \right) = \nu - c \quad (94)$$

$$\nu = \frac{\beta}{X_u} \left(\frac{1}{\beta} H_u(\rho) + c(X_- - X_+) - \langle \mathcal{E}_u(\rho) \rangle \right) \quad (95)$$

Where $X_+ = \sum_{s \in \mathcal{X}_+} x_s$, $X_- = \sum_{s \in \mathcal{X}_-} x_s$, $X_u = \sum_{s \in \mathcal{X}_u} x_s$, $H_u(\rho) = -\sum_{s \in \mathcal{X}_u} x_s \log x_s$ and $\langle \mathcal{E}_u(\rho) \rangle = \sum_{s \in \mathcal{X}_u} x_s \mathcal{E}_s$. (87)–(92) let us to relate the remaining Lagrange parameters by

$$\bar{\mu}_{i'} = f(x_{i'} - x_i e^{\beta(\mathcal{E}_i - \mathcal{E}_{i'})}) + \bar{\mu}_i e^{\beta(\mathcal{E}_i - \mathcal{E}_{i'})} \quad (96)$$

$$\mu_{j'} = f(x_j e^{\beta(\mathcal{E}_j - \mathcal{E}_{j'})} - x_{j'}) + \mu_j e^{\beta(\mathcal{E}_j - \mathcal{E}_{j'})} \quad (97)$$

$$(98)$$

$$\bar{\mu}_i = \frac{(k - \mu_j)(e^{\beta \mathcal{E}_i} x_i - e^{\beta(2c - \mathcal{E}_j)} x_j) - \mu_j e^{\beta(2c - \mathcal{E}_j)}}{(1 - x_i) e^{\beta \mathcal{E}_i} + x_j e^{\beta(2c - \mathcal{E}_j)}}$$

where $k = f - \bar{\mu}_i + \mu_j$. Summing (96) over i' and (97) over j' gives

$$\sum_{i'} \bar{\mu}_{i'} = f(X_+ - x_i e^{\beta \mathcal{E}_i} \mathcal{Z}_+) + \bar{\mu}_i e^{\beta \mathcal{E}_i} \mathcal{Z}_+ \quad (99)$$

$$\sum_{j'} \mu_{j'} = f(x_j e^{\beta \mathcal{E}_j} \mathcal{Z}_- - X_i) + \mu_j e^{\beta \mathcal{E}_j} \mathcal{Z}_- \quad (100)$$

where $\mathcal{Z}_+ = \sum_{s \in \mathcal{X}_+} e^{-\beta \mathcal{E}_s}$ and $\mathcal{Z}_- = \sum_{s \in \mathcal{X}_-} e^{-\beta \mathcal{E}_s}$. $f = 1 + \sum_{i'} \bar{\mu}_{i'} - \sum_{j'} \mu_{j'}$ therefore we can get f in terms of $\bar{\mu}_i$ and μ_j alone

$$f = \frac{1 + \bar{\mu}_i e^{\beta \mathcal{E}_i} \mathcal{Z}_+ - \mu_j e^{\beta \mathcal{E}_j} \mathcal{Z}_-}{x_i e^{\beta \mathcal{E}_i} \mathcal{Z}_+ + x_j e^{\beta \mathcal{E}_j} \mathcal{Z}_- + X_u} \quad (101)$$

We now have f in terms of $\bar{\mu}_i$ and μ_j , and (98) relates these to each other, so we can solve (93) and (94) simultaneously (using $k = f - \bar{\mu}_i + \mu_j$) to find $\bar{\mu}_i$ and μ_j , and by (96), (97) all Lagrange multipliers. Substituting (96) and (97) into (87) and solving for μ_j gives

$$\mu_j = \frac{k e^{\beta \mathcal{E}_i} (e^{-\beta c} - x_j e^{\beta \mathcal{E}_j + \nu})}{e^{\beta(\mathcal{E}_i - c)} + e^{\beta(\mathcal{E}_j + c)} + (1 - x_i - x_j) e^{\beta(\mathcal{E}_i + \mathcal{E}_j + \nu)}} \quad (102)$$

and similarly for $\bar{\mu}_i$

$$\bar{\mu}_i = \frac{ke^{\beta\mathcal{E}_i}(e^{\beta c} - x_i e^{\beta\mathcal{E}_i + \nu})}{e^{\beta(\mathcal{E}_i - c)} + e^{\beta(\mathcal{E}_j + c)} + (1 - x_i - x_j)e^{\beta(\mathcal{E}_i + \mathcal{E}_j) + \nu}} \quad (103)$$

Using (101) and $k = f - \mu_j + \bar{\mu}_i$ simultaneously solve (102) and (103) to give

$$\bar{\mu}_i = \frac{x_i e^\nu - e^{-\beta(\mathcal{E}_i - c)}}{e^\nu X_u + \mathcal{Z}_+ e^{\beta c} + \mathcal{Z}_- e^{-\beta c}} \quad (104)$$

$$\mu_j = \frac{e^{-\beta(\mathcal{E}_j + c)} - x_j e^\nu}{e^\nu X_u + \mathcal{Z}_+ e^{\beta c} + \mathcal{Z}_- e^{-\beta c}} \quad (105)$$

it is simple to check that (96) and 97 result in exactly the same equation for all $\bar{\mu}_{i'}$ and $\mu_{j'}$ but with the corresponding index. Now armed with the explicit form of the Lagrange parameters we have solved the Lagrangian. It is easy to check that for these solutions for $\bar{\mu}_i, \mu_j$ the Lagrangian simplifies to $\mathcal{L} = \sum_s x_s w_s$ where $w_s = 1/\beta \log(f_s e^{\beta\mathcal{E}_s} \mathcal{Z}')$ where the f_i are now of the form

$$f_s = \gamma^{-1} \begin{cases} x_s e^\nu, & s \in \mathcal{X}_u \\ e^{\beta(c - \mathcal{E}_s)}, & s \in \mathcal{X}_+ \\ e^{-\beta(c + \mathcal{E}_s)}, & s \in \mathcal{X}_- \end{cases} \quad (106)$$

where

$$\gamma = e^\nu X_u + \mathcal{Z}_+ e^{\beta c} + \mathcal{Z}_- e^{-\beta c} \quad (107)$$

Substituting in the above values for f_s into $\mathcal{L} = 1/\beta \sum_s x_s \log(f_s e^{\beta\mathcal{E}_s} \mathcal{Z}')$ gives our final result for work extraction

$$W^{(c)}(\rho) = \frac{1}{\beta} (\log \mathcal{Z}' - \log \gamma) \quad (108)$$

Next we show that γ can be written as $\gamma = \sum_s x_s^0 e^{-\beta\mathcal{E}_s} e^{\beta\theta_s}$ where θ_s is the fluctuation associates with subspace $|s\rangle\langle s|$ which, in the c -bounded distribution, is given by

$$\theta_s = \begin{cases} \frac{1}{\beta} \log(x_s e^{\beta\mathcal{E}_s}) + \frac{\nu}{\beta}, & s \in \mathcal{X}_u \\ +c, & s \in \mathcal{X}_+ \\ -c, & s \in \mathcal{X}_- \end{cases} \quad (109)$$

In general, the c -bounded work is given by the difference between the free energy of the final thermal state and the c -bounded free energy

$$F^{(c)}(\rho) = -\log \sum_s x_s^0 e^{-\beta\mathcal{E}_s} e^{\beta\theta_s} \quad (110)$$

B. Finding the optimal partition

In this appendix we derive the inequalities for partitioning the state space into positively bounded \mathcal{X}_+ , negatively bounded \mathcal{X}_- and unbounded \mathcal{X}_u energy levels, as required by our main result. We derive the general set of inequalities for any state and Hamiltonian and give worked through examples of how to find the partition for arbitrary 2 and 3 dimensional systems. Firstly we derive the partition inequalities for work extraction protocols,

and then show that the partition inequalities for work of formation are identical.

The Lagrangian is maximised under the condition that $\bar{\mu}_i$ and μ_j given in (102) and (103) are positive. Therefore if our optimization gives a negative Lagrange parameter it is set to zero, removing the corresponding c -bound. The Lagrange parameters are positive when the following inequalities are satisfied

$$\frac{1}{\beta} \log(x_i e^{\beta\mathcal{E}_i}) > c - \frac{\nu}{\beta} \rightarrow \bar{\mu}_i > 0 \quad (111)$$

$$\frac{1}{\beta} \log(x_j e^{\beta\mathcal{E}_j}) < -c - \frac{\nu}{\beta} \rightarrow \mu_j > 0 \quad (112)$$

where

$$\nu = \frac{\beta}{X_u} \left(\frac{1}{\beta} H_u(\rho) + c(X_- - X_+) - \langle \mathcal{E}_u(\rho) \rangle \right) \quad (113)$$

and we have used $\gamma \geq 0$. Clearly ν depends on how you partition the state space in to positively, negatively and unbounded fluctuations. In the following we derive a set of inequalities that determine the unique partition give ρ and c . The following observations simplify the problem

Lemma 2. *For any c -bounded work distribution where bounded fluctuations saturate their bounds, $X_\pm < 1/2$*

Proof. Consider a c -bounded work distribution with work marginals $\{w_s\}$, average work $W = \sum_s x_s w_s$, and all work marginals obey $|w_s - W| \leq c$. Break the work distribution up into work marginals that give positive fluctuations $w_i \geq W$ and negative fluctuations $w_j < W$. Writing the fluctuations as $\theta_i = w_i - W$ and $\theta_j = W - w_j$ and substituting into $W = \sum_s x_s w_s$ we get

$$\sum_i x_i \theta_i = \sum_j x_j \theta_j \quad (114)$$

which merely states that the average (non-absolute value) fluctuation of a random variable is zero, as is always the case. Take the case that $\sum_{s \in \mathcal{X}_+} x_s > \sum_{s \in \mathcal{X}_-} x_s$ and $X_+ > 1/2$. Clearly $\sum_{s \in \mathcal{X}_-} x_s < 1/2$ so $X_- < 1/2$. If $X_+ \geq 1/2$ then, as bounded fluctuations saturate their bounds, (114) gives the inequality

$$\sum_{s \in \mathcal{X}_-} x_s c < \sum_{s \in \mathcal{X}_+} x_s \theta_s \quad (115)$$

taking the factor of $\sum_{s \in \mathcal{X}_+} x_s$ to the other side we can write the inequality

$$c < \frac{1}{\sum_{s \in \mathcal{X}_+} x_s} \sum_{s \in \mathcal{X}_-} x_s \theta_s \quad (116)$$

The right hand side is a convex sum, and all $\theta_s \geq 0$, therefore at least one $\theta_s > c$ contradicting the fact that the work distribution is c -bounded. Therefore we must have $X_+ < 1/2$. A similar argument for the case $\sum_{s \in \mathcal{X}_+} x_s < \sum_{s \in \mathcal{X}_-} x_s$ gives that $X_- < 1/2$. \square

The second observation is that the inequalities (111) and (112) obey a β -ordering hierarchy. The β -ordered state, $\rho^{\downarrow\beta}$, is defined as

$$\rho^{\downarrow\beta} = (x_1, x_2, \dots, x_R) \quad , \quad x_s e^{\beta \mathcal{E}_s} \geq x_{s+1} e^{\beta \mathcal{E}_{s+1}} \quad (117)$$

where $R = \text{rank}(\rho)$ (there is no work marginal $w_s^{(\infty)}$ associated with $x_s = 0$). The β -ordering gives $w_i^{(\infty)} \geq w_{i+1}^{(\infty)}$. Therefore if $w_i^{(\infty)}$ satisfies (111) then so does $w_{i-1}^{(\infty)}$, and if $w_j^{(\infty)}$ satisfies (112) then so does $w_{j+1}^{(\infty)}$. From this we can deduce that the partition of the state space will look like

$$\rho^{\downarrow\beta} = (\underbrace{x_1, x_2, \dots, x_k}_{\mathcal{X}_+}, \underbrace{x_{k+1}, \dots, x_{l-1}}_{\mathcal{X}_u}, \underbrace{x_l, \dots, x_R}_{\mathcal{X}_-}) \quad (118)$$

Finally, it will be useful to put the inequalities in to the following form. In the following we drop the (∞) superscript from the unbounded work marginals $w_s^{(\infty)}$. Subscript i will label positively bounded fluctuations and subscript j negatively bounded fluctuations. (111) and (112) can be written in terms of the unbounded work distribution only (therefore the problem of finding the partition, assuming it exists, is a closed form)

$$W^{(\infty)}(\rho) < X_u(w_{i'} - c) + \sum_{s \in \mathcal{X}_+} x_s(w_s - c) + \sum_{s \in \mathcal{X}_-} x_s(w_s + c) \quad \forall i' \in \mathcal{X}_+ \quad (119)$$

$$W^{(\infty)}(\rho) > X_u(w_{j'} + c) + \sum_{s \in \mathcal{X}_+} x_s(w_s - c) + \sum_{s \in \mathcal{X}_-} x_s(w_s + c) \quad \forall j' \in \mathcal{X}_- \quad (120)$$

where $W^{(\infty)}(\rho)$ is the unbounded average work $W^{(\infty)}(\rho) = \beta^{-1} \log \mathcal{Z}' + F(\rho)$ and $X_u = 1 - X_+ - X_-$. We now prove that for a given ρ and c there exists a unique partition for which all inequalities (119) and (120) are satisfied.

Lemma 3. *Given state ρ , bound value c and inverse temperature β , there is a unique partition of the state space that gives the optimal c -bounded work. For $R = \text{rank}(\rho)$ there are at most $R - 1$ inequalities that must be checked to determine the partition.*

Proof. By 2 we know that $X_{\pm} < 1/2$, therefore $X_u > 0$. The partition obeys the β -ordering hierarchy (118). Starting from x_1 count left to right in $\rho^{\downarrow\beta}$ until you find the furthest x_k s.t. $\sum_{i=1}^k x_s < 1/2$. These will form our trial set for the bounded positive fluctuations $\tilde{\mathcal{X}}_+$. Starting from x_R count from right to left until you find the furthest x_l s.t. $\sum_{j=l}^R x_j < 1/2$. These form our trial set for the bounded negative fluctuations $\tilde{\mathcal{X}}_-$. If this is indeed the correct partition for the state space, the tightest bounds will be (119) on w_k and (120) on w_l .

The inequalities are

$$(1 - \sum_i x_i - \sum_j x_j)(w_k - c) + \sum_i x_i(w_i - c) + \sum_j x_j(w_j + c) > W(\rho) \quad (121)$$

$$(1 - \sum_i x_i - \sum_j x_j)(w_l + c) + \sum_i x_i(w_i - c) + \sum_j x_j(w_j + c) < W(\rho) \quad (122)$$

Whenever an inequality is satisfied it fixes that fluctuation as being the infinitum of its set. For example, if (121) is satisfied for w_k then $k \in \mathcal{X}_+$ regardless of \mathcal{X}_- . To see this, consider the case that (121) is satisfied for w_k but (122) isn't for w_l

$$X_u(w_k - c) + \sum_{i=1}^k x_i(w_i - c) + \sum_{j=l}^R x_j(w_j + c) > W(\rho)$$

$$X_u(w_l + c) + \sum_{i=1}^k x_i(w_i - c) + \sum_{j=l}^R x_j(w_j + c) > W(\rho)$$

as w_l doesn't satisfy its bound we have to move further down to $w_{l' > l}$. This makes $X_u \rightarrow X_u + \sum_{j=l}^{l'} x_j$ and $\sum_{j=l}^R x_j(w_j + c) \rightarrow \sum_{j=l}^R x_j(w_j + c) - \sum_{j=l}^{l'} x_j(w_j + c)$. The left hand side of (121) gains the term

$$+ \sum_{j=l}^{l'} x_j(w_k - c) - \sum_{j=l}^{l'} x_j(w_j + c) \quad (123)$$

$$= \sum_{j=l}^{l'} x_j(w_k - w_j)$$

As $w_k \geq w_j \quad \forall j = l, \dots, l'$ this term is positive and the new (121) is guaranteed to be satisfied. Therefore we see that there is a unique partition as satisfying an inequalities fixes the corresponding (positive or negative) subspace. There are at most $R - 1$ inequalities that we need to check, i.e. the “worst case” being when the state space is unbounded and we check all $R - 1$ inequalities. \square

In summary, the algorithm for determining the partition can be summarised as follows

1. β -order the state $\rho^{\downarrow\beta} = (x_1, \dots, x_R)$ where $x_s^{\beta \mathcal{E}_s} \geq x_{s+1}^{\beta \mathcal{E}_{s+1}}$ and $R = \text{rank}(\rho)$. This ordering gives the unbounded work distribution in descending order (w_1, \dots, w_R) where $w_s = \beta^{-1} \log(x_s e^{\beta \mathcal{E}_s} \mathcal{Z}')$ and $w_s \geq w_{s+1}$
2. Take the trial partition where you maximise $X_+ = \sum_{i=1}^k x_i$ under the condition $X_+ < 1/2$ and $X_- =$

$\sum_{j=l}^R x_j$, $X_- < 1/2$. Check inequalities

$$X_u(w_k - c) + \sum_{i=1}^k x_i(w_i - c) + \sum_{j=l}^R x_j(w_j + c) > W(\rho) \quad (124)$$

$$X_u(w_l + c) + \sum_{i=1}^k x_i(w_i - c) + \sum_{j=l}^R x_j(w_j + c) < W(\rho) \quad (125)$$

3. If (124) is satisfied x_k fixes $X_+ = \sum_{i=1}^k x_i$ and similar for (125). Otherwise we perform the next set of inequalities lower in the hierarchy, with $X_+ = \sum_{i=1}^{k-1} x_i$ if (124) is not satisfied and/or $\sum_{j=l+1}^R x_j$ if (125) is not satisfied. We repeat this process until we find a pair of inequalities that are simultaneously satisfied, fixing $\mathcal{X}_\pm = \mathcal{X}_\pm$, requiring at most $R - 1$ inequalities to be checked.

Case: $d=2$. This is the simplest case, as either $x_1 > 1/2$ or $x_2 > 1/2$ or they both $= 1/2$. In the first case we bound the negative fluctuation, and there is a single bound to check

$$w_2 < W^{(\infty)}(\rho) - c \quad (126)$$

and in the second case we bound the positive fluctuation if

$$w_1 > W^{(\infty)}(\rho) + c \quad (127)$$

and if $x_1 = x_2 = 1/2$ when the two fluctuations must be equal (as the average of the positive fluctuations = the average of the negative fluctuations) and we can choose to bound one or the other, giving the same free energy.

In the following section we show that the same algorithm is used for determining the partition for the c -bounded work of formation. Note that, unless all fluctuations are unbounded $W^{(c)}(\rho) < W^{(\infty)}(\rho)$ and, as shown in the next section, $W_F^{(c)}(\rho) > W_F^{(\infty)}(\rho)$.

IV. WORK OF FORMATION WITH BOUNDED FLUCTUATIONS

In this appendix we derive the c -bounded minimal work of formation $W_F^{(c)}$.

A. determining the optimal work of formation map

Assume there exists some optimal choice of $\{t(s'|s), w_{ss'}\}$ that minimize the work cost whilst obeying the c -bounds and microscopic reversibility (58). We start in the thermal state with $x_s = 1/Z e^{-\beta \mathcal{E}_s}$ and end the protocol in state ρ with probabilities $x_{s'}$. The average work is

$$W = \sum_{ss'} \frac{1}{Z} e^{-\beta \mathcal{E}_s} t(s'|s) w_{ss'} \quad (128)$$

where the t_{ij} must obey (55), (56) and (57).

Lemma 4. *We can always choose new protocol parameters $\tilde{t}_{s'} = x_{s'}$ and $\tilde{w}_{s'} = \sum_s \frac{t(s'|s) e^{-\beta \mathcal{E}_s}}{x_{s'} Z} w_{ss'}$ that give the same average work as the optimal protocol and obey all the necessary constraints*

Proof. Clearly these choices of $\tilde{t}_{s'}$ satisfy $\sum_{s'} \tilde{t}_{s'} = 1$ and $\sum_s 1/Z e^{-\beta \mathcal{E}_s} \tilde{t}_{s'} = x_{s'}$, and the average work is given by

$$\begin{aligned} \tilde{W} &= \sum_{ss'} 1/Z e^{-\beta \mathcal{E}_s} \tilde{t}_{s'} \tilde{w}_{s'} = \sum_{s'} \tilde{t}_{s'} \tilde{w}_{s'} \\ &= \sum_{ss'} x_{s'} \frac{t(s'|s) e^{-\beta \mathcal{E}_s}}{Z x_{s'}} w_{ss'} = W \end{aligned} \quad (129)$$

Note that $\sum_s \frac{t(s'|s) e^{-\beta \mathcal{E}_s}}{x_{s'} Z} = 1$ so $\tilde{w}_{s'}$ is a convex sum of the $w_{ss'}$. The reversibility inequality (58) demands that

$$\sum_s \tilde{t}_{s'} e^{\beta(\tilde{w}_{s'} - \Delta \mathcal{E}_{ss'})} \leq 1 \quad (130)$$

which we can see to be true as the LHS is

$$\begin{aligned} \text{LHS} &= x_{s'} e^{\beta(\tilde{w}_{s'} + \mathcal{E}_{s'})} \sum_s e^{-\beta \mathcal{E}_s} \\ &= x_{s'} Z e^{\beta(\tilde{w}_{s'} + \mathcal{E}_{s'})} \\ &= x_{s'} Z e^{\beta \mathcal{E}_{s'}} e^{\beta \sum_i \frac{t(s'|s) e^{-\beta \mathcal{E}_s}}{x_{s'} Z} w_{ss'}} \\ &\leq x_{s'} Z e^{\beta \mathcal{E}_{s'}} \sum_s \frac{t(s'|s) e^{-\beta \mathcal{E}_s}}{x_{s'} Z} e^{\beta w_{ss'}} \\ &= \sum_s t(s'|s) e^{\beta(w_{ss'} - \Delta \mathcal{E}_{ss'})} \leq 1 \end{aligned}$$

where in the fourth step we have used the convexity of the exponential function and in the last step we have used the fact that the optimal protocol defined by $\{t(s'|s), w_{ss'}\}$ is reversible.

Clearly the new protocol obeys the c -bounds as $\tilde{w}_{s'}$ is a convex combination of $w_{ss'}$ and the average of the work distribution is the same for both maps, therefore the spread of $\tilde{w}_{s'}$ about the average is less than or equal to that of $w_{ss'}$. \square

Substituting $t(s'|s) \rightarrow \tilde{t}_{s'} = x_{s'}$ and $w_{ss'} \rightarrow \tilde{w}_{s'}$ into the Lagrangian and simplifying gives

$$\mathcal{L} = \sum_{s'} f_{s'} w_{s'} + \sum_{s'} \lambda_{s'} (1 - x_{s'} e^{\beta \mathcal{E}_{s'}} Z e^{\beta w_{s'}}) + c \sum_{s'} (\mu_{s'} + \bar{\mu}_{s'}) \quad (131)$$

where we have dropped the tilde from $\tilde{w}_{s'}$ and $f_{s'} = x_{s'}(1 + f) - (\bar{\mu}_{s'} - \mu_{s'})$ and $f = \sum_{s'} (\bar{\mu}_{s'} - \mu_{s'})$. $\bar{\mu}$ give the bounds on positive fluctuations and μ give the bounds for negative fluctuations. Maximizing w.r.t $w_{s'}$ and $\lambda_{s'}$ gives

$$\lambda_{s'} = \max\left\{\frac{f_{s'}}{\beta}, 0\right\} \quad (132)$$

$$w_{s'} = -\frac{1}{\beta} \log\left(\frac{\beta \lambda_{s'} x_{s'} e^{\beta \mathcal{E}_{s'}} Z}{f_{s'}}\right) \quad (133)$$

$$= -\frac{1}{\beta} \log(x_{s'} e^{\beta \mathcal{E}_{s'}} Z), \quad f_{s'} \neq 0 \quad (134)$$

which is the work marginal given in the unbounded cases (it contains no $\bar{\mu}_j, \mu_j$ terms). Therefore all fluctuations

with $f_{s'} > 0$ are unbounded, with $\mu_{s'}(\bar{\mu}_{s'}) = 0$, giving $f_{s'} = x_{s'}$ if unbounded or 0 if bounded. The Lagrangian reduces to

$$\mathcal{L} = -\frac{1}{\beta} \sum_{s'} f_{s'} \log(x_{s'} e^{\beta \mathcal{E}_{s'}} \mathcal{Z}) + c \sum_{s'} (\bar{\mu}_{s'} + \mu_{s'}) \quad (135)$$

Which we can now re-write in terms of bounded and unbounded fluctuations. Index all fluctuations with $f_{s'} > 0$ with $s' \in \mathcal{X}_u$, and all those with $f_{s'} = 0$ with $s' \in \mathcal{X}_+$. The Lagrangian becomes

$$\mathcal{L} = -\frac{1}{\beta} \sum_{s' \in \mathcal{X}_u} x_{s'} (1+f) \log(x_{s'} e^{\beta \mathcal{E}_{s'}} \mathcal{Z}) + c \sum_{s' \in \mathcal{X}_+} (\bar{\mu}_s + \mu_s) \quad (136)$$

$f_s = 0 \forall s' \in \mathcal{X}_+$ gives

$$\bar{\mu}_{s'} - \mu_{s'} = x_{s'} (1+f) \quad (137)$$

Summing over $s \in \mathcal{X}_+$ and solving for $f = \sum_{k'} (\bar{\mu}_{k'} - \mu_{k'})$ gives

$$f = \frac{X_c}{1 - X_c} \quad (138)$$

where $X_c = \sum_{s' \in \mathcal{X}_+} x_{s'}$. For a given bounded fluctuation only one of the μ Lagrange parameters is non-zero, corresponding to if the fluctuation is positive or negative. Using the above result gives

$$\bar{\mu}_k = x_k (1 + \frac{X_c}{1 - X_c}) = \frac{x_k}{1 - X_c} \quad (139)$$

$$\mu_l = -\frac{x_l}{1 - X_c} \quad (140)$$

The Lagrange parameters are restricted to being positive, $\mu_s \geq 0$ and $\bar{\mu}_s \geq 0 \forall s$. Therefore as $\mu_s = -x_s/(1 - X_c)$ which is ≤ 0 , therefore all $\mu_s = 0$ and we never bound negative fluctuations.

Using $X_u = 1 - X_c$ we arrive at the c -bounded optimal work of formation

$$W_F^{(c)} = \frac{1}{X_u} \left(-\frac{1}{\beta} \sum_{s' \in \mathcal{X}_u} x_{s'} \log(x_{s'} e^{\beta \mathcal{E}_{s'}} \mathcal{Z}) + c X_c \right) \quad (141)$$

This is exactly the work cost when we calculate the c -bounded work of formation in the following way. Take the optimal unbounded work distribution for forming a state (w_1, \dots, w_d) with average $W^{(\infty)}(\rho)$. Equation (141) is given by the solution to the equation

$$\begin{aligned} W &= \sum_{s \in \mathcal{X}_u} x_s w_s + \sum_{s \in \mathcal{X}_+} x_s (W + c) \\ \therefore W (1 - \underbrace{\sum_{s \in \mathcal{X}_+} x_s}_{X_u}) &= \sum_{s \in \mathcal{X}_u} x_s w_s + c X_+ \\ \therefore W &= \frac{1}{X_u} \left(\sum_{s \in \mathcal{X}_u} x_s w_s + c X_+ \right) \end{aligned}$$

I.e we simply take the optimal unbounded work distribution and if the largest positive fluctuation breaks $w_i - W \leq c$ we replace it with $W' + c$, recalculating the average each time.

B. finding the optimal partition for state formation

We now show this algorithm for finding the partition is identical to the algorithm derived in Appendix 2 for work extraction, in the case of positive fluctuation bounds only.

Result 6. *The partition of the state space that gives the c -bounded work of formation is found using the algorithm described in Lemma 3 but bounding only positive fluctuations.*

Proof. Take the final state ρ' and β -order it

$$\rho'^{\downarrow \beta} = (x_1, \dots, x_R) \quad , \quad x_{s'} e^{\beta \mathcal{E}_{s'}} \geq x_{s'+1} e^{\beta \mathcal{E}_{s'+1}} \quad (142)$$

where $R = \text{rank}(\rho)$ (we can discount work marginals in the unbounded work distribution associated with $x_s = 0$ as they have no probability of being observed). The unbounded work distribution obeys the inverse β -ordering as $w_{s'} = -\beta^{-1} \log(x_{s'} e^{\beta \mathcal{E}_{s'}} \mathcal{Z})$, therefore $w_1 \leq w_2 \leq \dots \leq w_R$. For a correct partition where $X_+ = \sum_{s'=k}^R x_i$ we require that $w_{s'} \leq W_F(\rho')^{(c)} + c \forall s < k$. As with the c -bounded work extraction protocol the β -ordering puts these inequalities into a hierarchy, with the bound for w_{k-1} being the tightest. As all bounds must be satisfied we need only check the tightest one. For the aforementioned partition the tightest bound is

$$w_{k-1} \leq \frac{1}{\sum_{i=k}^R x_i} \left(W^{(\infty)} - \sum_{s'=k}^R x_{s'} w_{s'} + c \sum_{s'=k}^R x_{s'} \right) \quad (143)$$

which can be re-arranged to give

$$(1 - \sum_{s'=k}^R x_{s'}) w_{k-1} + \sum_{s'=k}^R x_{s'} w_{s'} \leq W^{(\infty)} + c \quad (144)$$

which is simply the inequality used for checking a partition of the state space in the c -bounded work extraction protocol, in the case that there are no negative fluctuations that saturate their bounds (124). It can be simplified further to

$$\sum_{s'=1}^{k-1} x_{s'} (w_{k-1} - w_{s'}) \leq c \quad (145)$$

So the partition is defined by the largest k s.t.

$$\sum_{s'=1}^k x_{s'} (w_k - w_{s'}) > c \quad (146)$$

$$\sum_{s'=1}^{k-1} x_{s'} (w_{k-1} - w_{s'}) \leq c \quad (147)$$

All w_s with $s \geq k$ are bounded and all with $s < k$ are unbounded. Once again we have at most $R - 1$ inequalities to check. I.e. starting from $k = 2$ we check (146) for $k = 2, 3, \dots$ \square

V. PROPERTIES OF C-BOUNDED WORK

A. recovering standard and single-shot regimes

Result 7. $\lim_{c \rightarrow 0} W^{(c)}(\rho) = W_{\min}(\rho)$

Proof. As $c \rightarrow 0$ the inequalities that determine if a fluctuation saturates its bound (111) and (112) become

$$\begin{aligned} \frac{1}{\beta} \log(x_s e^{\beta \mathcal{E}_s}) &\geq -\nu/\beta \rightarrow s \in \mathcal{X}_+ \\ \frac{1}{\beta} \log(x_s e^{\beta \mathcal{E}_s}) &\leq -\nu/\beta \rightarrow s \in \mathcal{X}_- \end{aligned}$$

All fluctuations satisfy one of these bounds, except in the case that $x_s = 0$ (i.e. there is no fluctuation associated with state $|s\rangle$), therefore $X_u \rightarrow 0$. $\gamma = X_u e^\nu + \mathcal{Z}_+ e^{\beta c} + \mathcal{Z}_- e^{-\beta c} \rightarrow \mathcal{Z}_+ + \mathcal{Z}_-$ which can be formulated as

$$\gamma = \sum_s x_s^0 e^{-\beta \mathcal{E}_s} \quad (148)$$

therefore $W^{(c)}(\rho)$ becomes

$$\begin{aligned} \lim_{c \rightarrow 0} W^{(c)}(\rho) &= \frac{1}{\beta} \left(\log \mathcal{Z}' - \log \sum_s x_s^0 e^{-\beta \mathcal{E}_s} \right) \\ &= W_{\min}(\rho) \end{aligned} \quad (149)$$

□

Result 8. $\lim_{c \rightarrow 0} W_F^{(c)}(\rho) = W_F^{\min}(\rho)$

Proof. Following the partition algorithm derived in Lemma 6, clearly when $c \rightarrow 0$ we bound all but the most negative fluctuation(s), those with value w_1 , as this gives the only positively bounded work distribution for which all work values are $\leq W_F^{(0)}$. Any other choice of partition would require both negative and positive bound saturation, which contradicts (139) being positive. Therefore $W_F^{(0)}$ is given by

$$W_F^{(0)} = \min_s \left\{ -\frac{1}{\beta} \log(x_s e^{\beta \mathcal{E}'_s} \mathcal{Z}) \right\} = -\frac{1}{\beta} \log(x_1 e^{\beta \mathcal{E}'_1} \mathcal{Z}) \quad (150)$$

where $\rho'^{\downarrow\beta} = (x_1, \dots, x_R)$. We can interpret

$$w_1 = -\frac{1}{\beta} \log(x_1 e^{\beta \mathcal{E}'_1} \mathcal{Z}) \quad (151)$$

$-w_1$ is the “on ramp” (the first segment) of the Lorenz curve of state ρ' (see Figure 4 and [1]). It therefore gives the work that can be extracted from the “thermally sharp state”, where all segments either have the same gradient or zero gradient, that just thermomajorizes ρ' , see figure 4 below and also [1]. This can also be interpreted as the upper bound to the work that can be in-deterministically extracted from ρ' , and this is precisely the single-shot work of formation. □

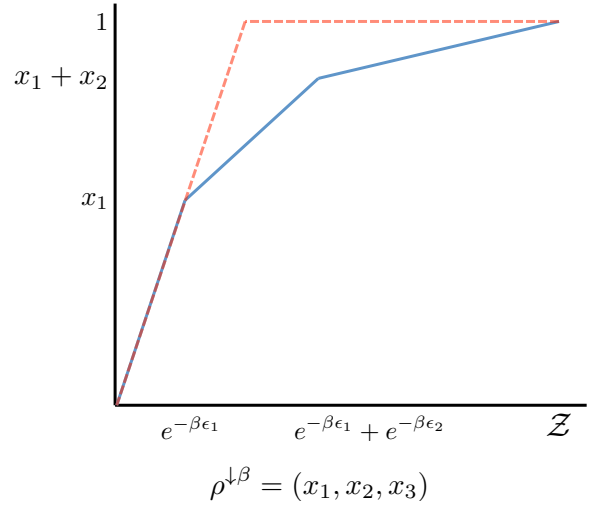


FIG. 4: Figure shows the thermally sharp state whose single-shot work content gives the work of formation of state ρ' . This also represents the maximal average work that can be extracted from ρ'

B. bounds on c-bounded work

Lemma 5. $W^{(c)}(\rho) \leq W_{\min}(\rho) + c$. $c \in [0, c_{\text{crit}}]$, therefore any gain in work extracted over the single-shot work content of ρ requires fluctuations that are at least as large as the increase in extracted work

Proof. For $c = 0$ we get that $W_{c=0}(\rho) = W_{\min}(\rho)$. $W^{(c)}$ increases as we increase c from 0. We show $W^{(c)}(\rho)$ grows sub-linearly with c , i.e. that

$$\frac{\partial W^{(c)}(\rho)}{\partial c} \leq 1 \quad \forall c \quad (152)$$

which implies that $|W^{(c)}(\rho) - W_{\min}(\rho)| \leq c$. Using (108) the derivative of $W^{(c)}(\rho)$ with respect to c is

$$\frac{\partial W^{(c)}(\rho)}{\partial c} = -\frac{1}{\beta \gamma} \frac{\partial \gamma}{\partial c} = \frac{1}{\gamma} ((X_+ - X_-) e^\nu - \mathcal{Z}_+ e^{\beta c} + \mathcal{Z}_- e^{-\beta c}) \quad (153)$$

where $\gamma = X_u e^\nu + \mathcal{Z}_+ e^{\beta c} + \mathcal{Z}_- e^{-\beta c}$. For this to be greater than 1 we require that

$$(X_+ - X_-) e^\nu - \mathcal{Z}_+ e^{\beta c} + \mathcal{Z}_- e^{-\beta c} > X_u e^\nu + \mathcal{Z}_+ e^{\beta c} + \mathcal{Z}_- e^{-\beta c} \quad (154)$$

As $\mathcal{Z}_\pm e^{\pm \beta c} \geq 0$ and $e^\nu \geq 0$ this cannot be satisfied unless

$$X_+ - X_- > X_u \dot{\Rightarrow} X_+ > \frac{1}{2} \quad (155)$$

where we have used $X_u = 1 - X_+ - X_-$. We now show that this is never true for c -bounded work distributions. Using $W^{(c)}(\rho) = \sum_s x_s \theta_s$, where $\theta_s = w_s - W^{(c)}(\rho)$ is the fluctuation associated with work marginal w_s , gives $\sum_s x_s \theta_s = 0$. Divide the fluctuations for the c -bounded protocol θ_s into positive $\theta_\alpha \geq 0$ where $\alpha \in \mathcal{X}_+$ and negative $\theta_\beta < 0$ where $\beta \in \mathcal{X}_-$, giving

$$\sum_\alpha x_\alpha |\theta_\alpha| = \sum_\beta x_\beta |\theta_\beta| \quad (156)$$

according to (85) and (86), bounded fluctuations saturate their bounds. Take the subset of the positive fluctuations that saturate their bounds as $\theta_{\alpha'}$. (156) becomes

$$cX_+ + \sum_{\alpha \neq \alpha'} x_\alpha |\theta_\alpha| = \sum_\beta x_\beta |\theta_\beta| \quad (157)$$

Dividing both sides by $\sum_\beta x_\beta$ gives

$$\frac{\sum_\beta x_\beta |\theta_\beta|}{\sum_\beta x_\beta} = \frac{cX_+ + \sum_{\alpha \neq \alpha'} x_\alpha |\theta_\alpha|}{\sum_\beta x_\beta} \quad (158)$$

the right hand side is strictly larger than c as $X_+ > 1/2$ and therefore $\sum_\beta x_\beta < 1/2$, and $\sum_{\alpha \neq \alpha'} x_\alpha |\theta_\alpha| \geq 0$. The left hand side is a convex sum with all $|\theta_\beta| \geq 0$, therefore at least one $|\theta_\beta|$ must be larger than c . Therefore in a c -bounded protocol $X_+ \leq 1/2$. \square

Lemma 6. $|W_F^{(c)}(\rho')| \geq |W_{\max}(\rho')| - c$. $c \in [0, c_{\text{crit}}]$, therefore any gain in work extracted over the single-shot work of formation of ρ' requires fluctuations that are at least as large as the increase in extracted work

Proof. if $|W_F^{(c)}(\rho')| < |W_{\max}(\rho')| - c$ then, taking $w_1 = \max\{w_{s'}\} = \max\{\beta^{-1} \log x_{s'} e^{\beta \epsilon_{s'}} \mathcal{Z}\} = |W_{\max}(\rho')|$ when we require that

$$\frac{1}{X_u} \left(\sum_{s' \in \mathcal{X}_u} w_{s'} + c(1 - X_u) \right) + c < w_1 \quad (159)$$

which requires that

$$c < \sum_{s' \in \mathcal{X}_u} (w_1 - w_{s'}) \quad (160)$$

which contradicts (145) for any partition \square

C. c -bounded work obeys the Jarzinski equality

In this appendix we show that all c -bounded work distributions obey the Jarzinski equality. The equality is given in the case that a system begins in the thermal state of an initial Hamiltonian H_i and the Hamiltonian is transformed to H_f , causing the state to evolve to a state that is, potentially, out of equilibrium with the final Hamiltonian. The equality is stated as

$$\langle e^{-\beta w} \rangle = e^{-\beta \Delta F} \quad (161)$$

where w is the work random variable (i.e. the work marginals associated with the protocol), with the convention that they are positive if the work is done on the system, and ΔF is the free energy difference between the equilibrium states of H_S and H'_S , given by

$$\Delta F = \frac{1}{\beta} \log \left(\frac{\mathcal{Z}}{\mathcal{Z}'} \right) \quad (162)$$

Result 9. Any protocol that saturates all reversibility constraints immediately satisfies the Jarzinski equality

Proof. If a protocol, characterised by $\{t(s'|s), w_{ss'}\}$, saturates all reversibility constraints then

$$\sum_s t(s'|s) e^{\beta(w_{ss'} - \epsilon_s + \epsilon_{s'})} = 1 \quad (163)$$

Dividing by \mathcal{Z} and re-arranging gives

$$\sum_s \frac{e^{-\beta \epsilon_s}}{\mathcal{Z}} t(s'|s) e^{\beta w_{ss'}} = \frac{e^{-\beta \epsilon_{s'}}}{\mathcal{Z}} \quad (164)$$

if we sum over s' , the LHS of this is equivalent to $\langle e^{\beta w_{ss'}} \rangle$ with the initial state being thermal w.r.t H_S , $x_s = \mathcal{Z}^{-1} e^{-\beta \epsilon_s}$, and the RHS becomes \mathcal{Z}'/\mathcal{Z} , which is equal to $e^{\beta \Delta F}$. Note that we use the convention of positive work for work extracted from the system, hence our $w_{ss'} = -w_s$ are the negatives of the work marginals given in the Jarzinski equality. \square

Result 10. The work distributions associated with $W^{(c)}(\rho)$ and $W_F^{(c)}(\rho)$ obey the Jarzinski equality.

First we cover the case of work extraction, $W^{(c)}(\rho)$. The work marginals are given by

$$w_s = \frac{1}{\beta} \log(f_s e^{\beta \epsilon_s} \mathcal{Z}') \quad (165)$$

and the $t(s'|s)$ are given by

$$t_{s'} = \frac{e^{-\beta \epsilon_{s'}}}{\mathcal{Z}'} \quad (166)$$

as derived in Appendix 2. Note that

$$\sum_s f_s = 1 \quad (167)$$

The LHS of the reversibility constraints are

$$\begin{aligned} &= \sum_s \frac{e^{-\beta \epsilon_{s'}}}{\mathcal{Z}'} e^{\beta(w_s - \epsilon_s + \epsilon_{s'})} \\ &= \sum_s \frac{1}{\mathcal{Z}'} e^{\log(f_s \mathcal{Z}')} \\ &= \sum_s f_s = 1 \end{aligned}$$

and can similarly be shown for $W^{(c)}(\rho)$.

VI. QUBIT CARNOT ENGINE

In this appendix we calculate the c -bounded Carnot efficiency for a single qubit quantum engine. The engine operates by moving a qubit $\rho = x|0\rangle\langle 0| + (1-x)|1\rangle\langle 1|$ with Hamiltonian $H_s = \mathcal{E}|0\rangle\langle 0|$ between two baths with inverse temperature β_H and β_C , with $\beta_H < \beta_C$. The engine cycle begins with the qubit in thermal equilibrium with the cold bath. It is then placed in contact with the hot bath, and extracting work by allowing the state to

equilibrate. The final step in the cycle is to return the qubit to the cold bath, allowing it to equilibrate and extracting work.

Case: $c \rightarrow \infty$. Following the proof given in [2], we show that in the case that fluctuations are unbounded it is possible to reach Carnot efficiency. The equilibrium state of the qubit is given by

$$\rho_{H,C} = \frac{1}{\mathcal{Z}_{H,C}} e^{-\beta_{H,C}\mathcal{E}} |0\rangle\langle 0| + \frac{1}{\mathcal{Z}_{H,C}} |1\rangle\langle 1| \quad (168)$$

where $\mathcal{Z}_{H,C} = e^{-\beta_{H,C}\mathcal{E}} + 1$. When ρ_C is placed in thermal contact with the hot bath we can extract a maximal average work given by the free energy difference

$$F(\rho_C, \beta_H) - F(\rho_H, \beta_H) = \frac{1}{\beta_H} \log \left(\frac{\mathcal{Z}_H}{\mathcal{Z}_C} \right) + \text{tr}(H_s \rho_C) \left(\frac{\beta_H - \beta_C}{\beta_H} \right) \quad (169)$$

where $\text{tr}(H_s \rho_C) = \mathcal{Z}_C^{-1} e^{-\beta_C \mathcal{E}} \mathcal{E}$ and work marginals

$$\begin{aligned} w_{0,H} &= \frac{1}{\beta_H} \log \left(\frac{\mathcal{Z}_H}{\mathcal{Z}_C} e^{\mathcal{E}(\beta_H - \beta_C)} \right) \\ w_{1,H} &= \frac{1}{\beta_H} \log \left(\frac{\mathcal{Z}_H}{\mathcal{Z}_C} \right) \end{aligned} \quad (170)$$

where we have used the result for the optimal unbounded work extraction protocol $w_s = \beta^{-1} \log x_s e^{\beta \mathcal{E}_s} \mathcal{Z}$. Returning the qubit to the cold bath we extract work

$$F(\rho_H, \beta_C) - F(\rho_C, \beta_C) = \frac{1}{\beta_C} \log \left(\frac{\mathcal{Z}_C}{\mathcal{Z}_H} \right) + \text{tr}(H_s \rho_H) \left(\frac{\beta_C - \beta_H}{\beta_C} \right) \quad (171)$$

with work marginals

$$\begin{aligned} w_{0,C} &= \frac{1}{\beta_C} \log \left(\frac{\mathcal{Z}_C}{\mathcal{Z}_H} e^{\mathcal{E}(\beta_C - \beta_H)} \right) \\ w_{1,C} &= \frac{1}{\beta_C} \log \left(\frac{\mathcal{Z}_C}{\mathcal{Z}_H} \right) \end{aligned} \quad (172)$$

where $\text{tr}(H_s \rho_H) = \mathcal{Z}_H^{-1} e^{-\beta_H \mathcal{E}} \mathcal{E}$. The total work extracted in one cycle is

$$W_{\text{tot}} = \left(\frac{1}{\beta_H} - \frac{1}{\beta_C} \right) (S_H - S_C) \quad (173)$$

where $S_{H,C} = -\log \mathcal{Z}_{H,C} - \beta_{H,C} \text{tr}(H_s \rho_{H,C})$ is the Von Neumann entropy of the Gibbs state with inverse temperature $\beta_{H,C}$. Applying the first law of thermodynamics $\Delta U = Q - W$ to the first step (extracting work from the hot bath), we get

$$\text{tr}(H_s \rho_H) - \text{tr}(H_s \rho_H) = Q_H - (F(\rho_C, \beta_H) - F(\rho_H, \beta_H)) \quad (174)$$

which simplifies to

$$Q_H = \frac{1}{\beta_H} (S_H - S_C) \quad (175)$$

where Q_H is the heat flow out of the hot bath. The efficiency is given by the ratio of the total work to Q_H , giving

$$\frac{W_{\text{tot}}}{Q_H} = 1 - \frac{\beta_H}{\beta_C} \quad (176)$$

which is the Carnot efficiency.

Case: c finite.

$\mathcal{E} > 0$ implies that $x < 1/2$ for all thermal states of ρ , therefore by the partitioning algorithm derived in Appendix 4, the least likely fluctuation $w_0^{(c)}$ associated with subspace $|0\rangle$ is the only fluctuation that can be bounded. Using equation (16), the c -bounded work content for a qubit $\rho^{\downarrow \beta} = (x_1, x_2)$ is given by

$$W^{(c)}(\rho) = \frac{1}{\beta} \log \mathcal{Z} + \begin{cases} F(\rho), & x_1 e^{\beta \mathcal{E}_1} \leq e^{\beta(F(\rho) + c)} \\ & x_2 e^{\beta \mathcal{E}_2} \geq e^{\beta(F(\rho) - c)} \\ F_1^{(c)}(\rho), & w_0 < W(\rho) - c \\ F_2^{(c)}(\rho), & w_0 > W(\rho) + c \end{cases} \quad (177)$$

where $F_1^{(c)}(\rho) = -\beta^{-1} \log \left(e^{-c\beta + \frac{c\beta}{1-x}} + e^{-c\beta - \mathcal{E}\beta} \right)$ and $F_2^{(c)}(\rho) = -\beta^{-1} \log \left(e^{c\beta - \frac{c\beta}{1-x}} + e^{c\beta - \mathcal{E}\beta} \right)$, $W(\rho)$ is the unbounded work given by the free energy, and $w_0 = \beta^{-1} \log(x e^{\beta \mathcal{E}} \mathcal{Z})$. In the first step, putting the qubit in contact with the hot bath, the work we extract is bounded negatively if $w_{0,H} < F(\rho_C, \beta_H) - F(\rho_H, \beta_H) - c$, which gives the inequality

$$\frac{\mathcal{E}}{\mathcal{Z}_C} \left(\frac{\beta_C - \beta_H}{\beta_H} \right) > c \quad (178)$$

note that $\beta_C > \beta_H$ therefore the LHS of (178) is positive. In order to break the positive fluctuation bound we would require $-LHS > c$ which is impossible for positive c , therefore in the first part of the engine cycle the work extracted is either unbounded or negatively bounded. For the second cycle, re-equilibrating the qubit with the cold bath, the inequality that implies a positively bounded work is given by $w_{0,C} > F(\rho_H, \beta_C) - F(\rho_C, \beta_C) + c$, which simplifies to

$$\frac{\mathcal{E}}{\mathcal{Z}_H} \left(\frac{\beta_C - \beta_H}{\beta_C} \right) > c \quad (179)$$

Again the LHS is positive, and we would require $-LHS > c$ in the case of a negatively bounded protocol. Therefore on the second part of the cycle the work is either unbounded or positively bounded. Also note that as $\beta_H < \beta_C$ and therefore $\mathcal{Z}_C < \mathcal{Z}_H$, satisfying inequality (178) implies inequality (179) is also satisfied. There are therefore three cases, 1) the work is unbounded and we can achieve Carnot efficiency, 2) the work extracted from the hot bath is negatively bounded, and 3) 2 and the work extracted from the cold bath is positively bounded.

the c -bounded work extracted from the hot bath, in the case that inequality (178) is satisfied, is given by

$$W_1^{(c)} = \frac{1}{\beta_H} \log \left(\frac{\mathcal{Z}_H}{e^{c\beta_H \mathcal{Z}_C} + e^{-\beta_H \mathcal{E}}} \right) + c \quad (180)$$

where we have used the expressions in (177). If inequality (179) is satisfied for the second part of the cycle, then we get

$$W_2^{(c)} = \frac{1}{\beta_C} \log \left(\frac{Z_C}{e^{-c\beta_C Z_H} + e^{-\beta_C \mathcal{E}}} \right) - c \quad (181)$$

Using $Q_H = \Delta U + W_1^{(c)}$, the efficiency is given by

$$\eta_1^{(c)} = \frac{W_1^{(c)} + F(\rho_H, \beta_C) - F(\rho_C, \beta_C)}{\Delta U + W_1^{(c)}} \quad (182)$$

if inequality (178) is satisfied and

$$\eta_2^{(c)} = \frac{W_1^{(c)} + W_2^{(c)}}{\Delta U + W_1^{(c)}} \quad (183)$$

if inequalities (178) and (179) are satisfied. As the temperature difference between hot and cold baths increases, the efficiency increases. In the unbounded case, the Carnot efficiency is bounded from above by 1. As $\beta_H/\beta_C \rightarrow 0$, inequality (179) cannot be satisfied for any finite c . To find the upper bound of $\eta_1^{(c)}$ we make the change of variables $\beta_C/n = \beta_H$, i.e. $T_H = nT_C$ and n is the ratio of the two bath temperatures. We then

take $\beta_H \rightarrow \infty$ and $n \rightarrow \infty$.

$$\lim_{n \rightarrow 0} \left(\lim_{\beta_H \rightarrow \infty} \left(\eta_1^{(c)} \right) \right) = 1 - \frac{\mathcal{E}}{2(2c + \mathcal{E})} \quad (184)$$

Of course as $\mathcal{E} \rightarrow 0$ the efficiency tends to 1, but this is because we are no longer extracting any work (as the two Gibbs states are identical). As $c \rightarrow \infty$ we recover the Carnot efficiency upper bound, although for any finite c we can never reach an efficiency of 1. Therefore all realistic engines of this form have an upper bound to their efficiency defined by their energy gap and fragility c . This represents a general physical upper bound on the efficiency of a the qubit engine, given by its ability to withstand fluctuations. Also notice that as $c \rightarrow 0$ the maximal efficiency is bounded from below by $1/2$. For $c = 0$ the efficiency is zero, as no work can be extracted. Therefore we observe, in this model at least, a genuine discontinuity between the capabilities of a thermal machine in the single-shot regime. Allowing for arbitrarily small fluctuations in principle will allow the engine to reach a maximum efficiency that is bounded from below by $1/2$. But if we demand that the work is deterministic, the efficiency is zero.